

**Observatório Nacional**  
**Ministério da Ciência, Tecnologia e Inovação**  
**Programa de Pós-Graduação em Astronomia**

Dissertação de Mestrado

**Estudo de QSOs no levantamento Dark Energy Survey**

**Fernanda Ostrovski Dutra**

Dissertação de Mestrado apresentada no  
Observatório Nacional como parte dos requisitos  
para a obtenção do grau de Mestre em Astronomia

Orientador: Marcio Antonio Geimba Maia

Rio de Janeiro, abril de 2013

*Ao Pedro*

# Agradecimentos

Primeiramente gostaria de agradecer a meus pais, Henrique e Stella pelo amor e apoio incondicionais ao longo de toda a minha vida. Eles são os pilares que me permitiram buscar a carreira acadêmica e seguir meus sonhos. Sem eles, a conclusão de meu mestrado não seria possível.

Do mesmo modo, agradeço ao restante de minha família, em especial a minha irmã Juliana, meu cunhado Juan e minha avó Elisa, que mesmo longe, sempre torceu pelo meu sucesso e felicidade. Também não posso esquecer-me da Dona Sandra e da ajuda imensurável que ela me forneceu.

Agradeço ao meu orientador, Dr Marcio Maia, pelo acompanhamento e apoio na realização desse projeto e pelo auxílio no amadurecimento de meus conhecimentos e conceitos em astronomia que me fizeram crescer como aluna e futura pesquisadora.

Também não posso deixar de agradecer ao Dr Luiz Nicolaci pelo incentivo e pela franqueza de nossas conversas, ao Dr Paulo Pellegrini pelo estímulo acadêmico e pelas discussões científicas, aos colegas de trabalho Aurélio, Flávia e Ogando, sempre dispostos a oferecer ajuda e me salvar de problemas que pareciam irresolúveis, e à Rosane, sempre lá para um abraço.

Ao grupo de TI, Carlos, Cristiano, Wagner, Peloso, Ana Marcela, Rick, Patricia e Angelo, que sempre esteve disposto a auxiliar com todo tipo de problema computacional, pertinente ou não ao projeto e que nunca riu das minhas perguntas bobas ou dos meus erros de programação.

A todos os meus professores e colegas de disciplinas no mestrado com quem pude desfrutar o prazer de aprender e dar boas risadas em viagens de campo.

Aos meus amigos que nunca deixaram de me chamar para sair mesmo sabendo que nem sempre eu iria. Ao Rick, Milinha, Gui e Chico, sempre dispostos a mais uma partida de Detetive e uma fornada de pão de queijo. À Caroline e à Leticia com nossos dias de menina, e à Carol, sempre a um WhatsApp de distância.

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq, pela bolsa concedida durante os anos do curso. Ao LIneA e ao DES-Brazil por toda a estrutura que me foi dada para desenvolver este trabalho e pelas oportunidades científicas de que pude desfrutar ao participar de um projeto de colaboração internacional que me permitiu trabalhar com parceiros de instituições de primeira

linha no mundo. Agradeço a eles também pelo apoio financeiro em viagens internacionais. E agradeço à pós-graduação do Observatório Nacional pelo apoio financeiro em viagens no território brasileiro.

Finalmente, agradeço a meu marido Pedro por estar ao meu lado ao longo de mais uma etapa da minha vida. Obrigada por todos os almoços quentinhos que me esperaram, por todo o suporte, pela companhia constante seja nas derrotas ou nas vitórias e principalmente por todo o amor.

# Resumo

Quasares são objetos primordiais de grande importância para o estudo da evolução das galáxias e do Universo. Os grandes levantamentos que estão surgindo e planejam cobrir grandes áreas do céu a limites profundos de magnitude, irão permitir um aumento da amostra desses objetos, desenvolvendo nosso conhecimento sobre eles, e estenderão a fronteira do Universo observável com a descoberta de quasares a *redshifts* ( $z$ ) cada vez mais distantes. Um exemplo desse tipo de levantamento é o *Dark Energy Survey* (DES), que irá obter fotometria nas bandas  $g$ ,  $r$ ,  $i$ ,  $z$ , e  $Y$  em uma área de 5000 graus quadrados de céu. Dados os limites de magnitude do DES, esperamos aumentar a amostra de quasares na faixa de  $z$  entre 4 e 6, e também encontrar novos objetos acima de 7.

O objetivo desse trabalho é simular uma amostra de quasares similares aos que serão observados pelo DES, fazendo previsões sobre suas distribuições numéricas, de cor e de  $z$ . Com a geração de catálogos simulados, ferramentas de análise para a seleção de QSOs e cálculo de *redshifts* fotométricos (*photo- $z$* ) podem ser testadas antes de suas aplicações em dados reais. Para realizar a simulação, desenvolvemos uma suíte de algoritmos chamada AddQSO, que utiliza as facilidades do Portal Científico do DES, entre elas a infraestrutura e a apresentação global dos resultados, onde foi implementado.

O AddQSO integra as funções de luminosidade de quasares de Richards et al. (2006) para simular objetos de  $z < 5,25$  e de Willott et al. (2010) para o restante da faixa de  $z$ , o que nos fornece, além da distribuição em  $z$  dos quasares, a distribuição em magnitude absoluta na banda  $i$  e, conseqüentemente, a distribuição de magnitude aparente nessa mesma banda. Através da convolução entre distribuições espectrais de energia e a curva de transmissão dos filtros, as magnitudes aparentes nas outras bandas são simuladas. Erros fotométricos são acrescentados seguindo as distribuições desse parâmetro em uma simulação de estrelas realizada pela colaboração DES. A suíte foi validada através de uma comparação com dados do Sloan Digital Sky Survey, que mostra conformidade na distribuição de cores, evolução de cores com *redshift*, e número de quasares gerados.

Utilizamos os resultados do AddQSO para testar dois códigos de *photo- $z$* , ANNz e LePhare. O ANNz mostrou melhores resultados, apresentando apenas  $\approx 12\%$

de identificações catastróficas. O LePhare também foi testado como método de classificação de objetos com catálogos simulados de quasares, estrelas e galáxias. Apesar da completeza do método ser de 0,92, a pureza é de apenas 0,06.

Concluimos que o AddQSO é uma ferramenta robusta para simular catálogos de quasares. Para o levantamento DES, prevemos a detecção de  $10^7$  quasares em pelo menos uma de suas cinco bandas, sendo quase 800 objetos com  $z > 6,0$  e 20 com  $z > 7,0$ . O código ANNz apresenta resultados satisfatórios para o cálculo de *photo-z* e não parece ser afetado pela ausência da banda *u* como acontece com o algoritmo LePhare. Finalmente, concluimos que apenas uma pequena fração dos quasares pode ser separada através da aplicação de limites de cor e que a seleção de objetos apenas pelo método de ajuste de modelos não é suficiente para criar uma amostra de quasares devido à grande taxa de contaminação por outros tipos de objetos.

**Palavras-chave:** quasares, função de luminosidade, simulação de catálogos

# Abstract

Quasars are important primordial objects in the study of galaxy evolution and the Universe. The next surveys that will cover large sky areas to deep magnitude limits will allow the increase of the QSO sample, aiding in the development of our knowledge of these objects and extending the boundaries of the observable Universe through the discovery of objects at higher *redshifts* ( $z$ ). Amongst these surveys is the *Dark Energy Survey* (DES), which will observe 5000 square degrees using the  $g$ ,  $r$ ,  $i$ ,  $z$ , and  $Y$  filters. Given the DES magnitude limits, we expect to increase the sample of QSOs in the 4-6 *redshift* range in addition to finding new objects with  $z > 7.0$ .

The objective of this work is to simulate a mock quasar sample similar to the objects we expect to observe with DES, in order to predict number counts as well as color and  $z$  distributions. With mock catalogs, it is possible to test analysis tools such as QSO selection and photometric *redshifts* (*photo-z*) tools to be used with the real data once available. To simulate the objects, we have developed a suite of algorithms named AddQSO that has been implemented on the DES Science Portal and utilizes its infrastructure and facilities.

AddQSO integrates the quasar luminosity functions of Richards et al. (2006) to simulate objects with  $z < 5.25$  and of Willott et al. (2010) for the higher  $z$  range, which provides, in addition to the quasars  $z$  distribution, the distributions of absolute magnitude in the  $i$  band and, consequently, of the apparent magnitude in this band. Through the convolution of filter transmission curves and spectral energy distributions, the apparent magnitudes in other bands are simulated. Photometric errors are added following the distributions of this parameter in a stellar simulation performed by the DES collaboration. The suite was validated through a comparison to Sloan Digital Sky Survey data, which show conformity in the color distribution as well as the color evolution with *redshift* and the number of generated objects.

The AddQSO results were used for testing two *photo-z* codes, ANNz and LePhare. The former showed the best results, with only  $\approx 12\%$  catastrophic identifications. The latter was also tested as an object classification method using mock stars, galaxies and quasars catalogs. Though the completeness of the method was found to be of 0.92, the purity is only of 0.06.

We have concluded that AddQSO is a robust tool to simulate quasar catalogs. For the DES survey, we predict the existence of  $10^7$  quasars with photometry in

at least one of the five bands, of which 1000 objects have  $z > 6.0$  and 40 show  $z > 7.0$ . The ANNz code shows satisfactory results for *photo-z* measurements and does not seem to be affected by the lack of the *u* band as it occurs with the LePhare algorithm. Finally, we concluded that only a small fraction of QSOs can be selected through the definition of color cuts and that the object selection based solely on a template fitting method is not sufficient for the creation of a quasar sample due to the large contamination rates by other types of objects.

**Keywords:** quasars, luminosity function, mock catalogs



# Índice

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>O <i>pipeline</i> AddQSO</b>	<b>13</b>
2.1	Geração das distribuições de $z$ , $M_i$ e $m_i$ . . . . .	15
2.2	Geração de magnitudes aparentes em outros filtros . . . . .	22
2.3	Atribuição de erros para as magnitudes . . . . .	29
2.4	Distribuição espacial . . . . .	32
2.5	Validação dos algoritmos . . . . .	32
2.6	Previsões para o DES . . . . .	38
<b>3</b>	<b>Redshifts Fotométricos</b>	<b>44</b>
3.1	Redes Neurais . . . . .	45
3.2	Ajuste de modelos . . . . .	50
<b>4</b>	<b>Classificação de Objetos</b>	<b>55</b>
4.1	Classificação por Cor e Dropouts . . . . .	55
4.2	Seleção por Ajuste de Modelos . . . . .	57
<b>5</b>	<b>Aplicação das ferramentas de análise ao DES</b>	<b>63</b>
5.1	Os catálogos . . . . .	63
5.2	Classificação . . . . .	64
5.2.1	Seleção por cor e <i>dropouts</i> . . . . .	65
5.2.2	Classificação por ajuste de modelos . . . . .	69
5.3	Photo- $z$ . . . . .	71
5.3.1	Redes Neurais . . . . .	71
5.3.2	Ajuste de Modelos . . . . .	73
<b>6</b>	<b>Conclusões e Perspectivas</b>	<b>76</b>
<b>A</b>	<b>Códigos XML do AddQSO</b>	<b>80</b>
A.1	Estrutura do <i>workflow</i> do AddQSO . . . . .	80
A.2	Configuração do módulo de integração da QLF . . . . .	81

<b>B Atribuição de erros</b>	<b>84</b>
B.1 A adição de $\Delta m$ às bandas $g, r, z$ , e $Y$ . . . . .	84
B.2 A adição de incertezas às bandas $g, r, z$ , e $Y$ . . . . .	86
<b>Referências</b>	<b>88</b>

# Lista de Figuras

1.1	Espectro do quasar 3C273 em comparação com o espectro de uma galáxia elíptica não-ativa de tipo morfológico cD. A emissão da radiação do QSO é observada em quase toda a extensão do espectro eletromagnético com a energia por intervalo de frequência logarítmica sendo quase constante (Schneider, 2006). . . . .	4
1.2	O modelo unificado de núcleos ativos de galáxias. Ao redor do buraco negro supermassivo está orbitando um disco de acreção de gás e poeira. Frequentemente, há a presença de jatos de partículas alinhados com o eixo de rotação do buraco negro. O tipo de AGN a ser observado da Terra depende do ângulo de visada (crédito da imagem: NASA apud Pierre Auger Observatory - <a href="http://www.auger.org/news/PRagn/about_AGN.html">http://www.auger.org/news/PRagn/about_AGN.html</a> ). . . . .	7
2.1	Estrutura do <i>workflow</i> do <i>pipeline</i> AddQSO integrado no Portal Científico do DES. . . . .	14
2.2	Distribuições de <i>redshift</i> (painel da esquerda) e de $m_i$ (painel da direita) para os $\approx 15000$ quasares da amostra utilizada por Richards et al (2006) na determinação da QLF. . . . .	15
2.3	QLF de Richards et al. (2006) para diferentes intervalos de <i>redshift</i> . A função sendo usada está representada pela linha azul ponto-tracejada. A linha cinza representa a curva a $z = 2,01$ e foi reproduzida para comparação (Richards et al. 2006). Os números em cada painel mostram o centro dos intervalos de <i>redshift</i> , que têm 0,5 de largura. . . . .	17
2.4	QLF de Willott et al (2010) para $z = 6$ . A função sendo usada está representada pela linha preta contínua e é determinada em função de $M_{1450}$ . . . . .	19
2.5	Interface do Portal Científico mostrando a aba de seleção da QLF a ser integrada pelo AddQSO. . . . .	20
2.6	Parâmetros cosmológicos fiduciais na aba de configurações da interface AddQSO no Portal Científico. . . . .	20

2.7	Parâmetros fiduciais relativos à intergração da QLF pelo AddQSO na aba de configurações da interface do Portal Científico. . . . .	21
2.8	Curvas de transmissão dos filtros DES (esquerda) e SDSS (direita) usadas pelo AddQSO. . . . .	23
2.9	Interface do Portal Científico mostrando a aba de seleção do conjunto de filtros e das bibliotecas de SEDs a serem utilizadas pelo AddQSO na geração das magnitudes aparentes em outros filtros. . . . .	23
2.10	SEDs presentes na biblioteca de Budavári et al. (2001). As linhas contínuas e tracejadas distinguem os dois conjuntos de SEDs presentes na biblioteca. . . . .	25
2.11	SEDs presentes na biblioteca de Mari Polletta (SWIRE). Os modelos representados pelas linhas tracejadas são de quasares tipo 2. . . . .	25
2.12	Subamostra das SEDs presentes na biblioteca de espectros sintéticos de quasares. . . . .	26
2.13	Fator de transmissão de acordo com o <i>redshift</i> . Os pontos são dados em Songaila (2004) e o ajuste da curva é dado pela equação (2.8). . . . .	27
2.14	Distribuição de $\Delta m_i$ em função da magnitude simulada na banda $i$ . No painel da esquerda vemos em azul a distribuição de $\Delta m_i$ para as estrelas do DC6B e, sobrepostas em vermelho, as médias com barras de $1\sigma$ calculadas após a remoção dos objetos que apresentavam $\Delta m_i$ fora do limite de $3\sigma$ . No painel da direita está a distribuição de $\Delta m_i$ simulada para quasares em uma área de 5000 graus quadrados no filtro $i$ do DES. A barra colorida representa o logartimo da contagem de objetos em cada ponto. . . . .	31
2.15	Erros fotométricos na banda $i$ em função da magnitude observada na mesma banda. No painel da esquerda vemos em azul a distribuição de erro para as estrelas do DC6B e, sobrepostas em vermelho, as medianas e as barras de $1\sigma$ . No painel da direita está a distribuição de erros fotométricos simulada para quasares em uma área de 5000 graus quadrados no filtro $i$ do DES. A barra colorida representa o logartimo da contagem de objetos em cada ponto. . . . .	31
2.16	Distribuições de <i>redshift</i> para os $\approx 105000$ quasares do catálogo QDR7 (Schneider et al. 2010). . . . .	33
2.17	Comparação entre os diagramas cor-cor dos quasares do catálogo do QDR7 (cinza) e dos quasares simulados pelo AddQSO. As diferentes bibliotecas de SEDs estão codificadas pelas cores indicadas nos painéis. . . . .	34

2.18	Amostragem de SEDs de cada uma das bibliotecas presentes no AddQSO utilizadas na geração das magnitudes aparentes do catálogo SDSS_8417. As SEDs foram normalizadas em 3000 Å. As cores representam o mesmo que na figura 2.17. . . . .	35
2.19	Cores médias por intervalo de 1,0 em <i>redshift</i> . As cores representam o centro de cada intervalo. Os círculos mostram os valores para os quasares do catálogo SDSS_8417 e os triângulos denotam as médias do catálogo de quasares QDR7. . . . .	37
2.20	Contagens na banda <i>i</i> para o catálogo QR09 de $\approx 1 \times 10^6$ candidatos a quasares do SDSS (azul) e o catálogo de quasares simulados SDSS_8417 (vermelho). . . . .	38
2.21	Distribuição de <i>redshifts</i> dos quasares do catálogo DES_5000 sem cortes em magnitude. As linhas vermelhas representam as tendências de cada QLF usada na simulação caso elas fossem estendidas muito além de seus regimes de validade em <i>redshift</i> . . . . .	39
2.22	Distribuição de <i>redshift</i> dos quasares do catálogo DES_5000 que contém pelo menos uma das magnitudes dentro dos limites $m_g < 25, 21$ , $m_r < 24, 77$ , $m_i < 24, 04$ , $m_z < 23, 44$ , $m_Y < 21, 70$ . . . . .	40
2.23	Contagens esperadas em cada banda, dado os limites fotométricos $m_g < 25, 21$ , $m_r < 24, 77$ , $m_i < 24, 04$ , $m_z < 23, 44$ , $m_Y < 21, 70$ . . . . .	41
2.24	Diagramas cor-cor para os quasares do catálogo DES_5000. A barra colorida indica o <i>redshift</i> de cada objeto. . . . .	42
2.25	Diagramas de cor média por <i>redshift</i> dos quasares do catálogo DES_5000. . . . .	43
3.1	Diagrama esquemático de uma ANN com os neurônios da camada de entrada recebendo valores de magnitude em diferentes filtros e uma única camada oculta. O resultado é um neurônio na camada de saída contendo o photo-z (Firth et al. 2003). . . . .	46
3.2	Comparação entre os $z_{spec}$ do catálogo de quasares QDR7 e os $z_{photo}$ calculados pelo ANNz utilizando todos os cinco filtros SDSS. A escala de cor mostra a quantidade de objetos em cada ponto. . . . .	49
3.3	Comparação entre os $z_{spec}$ do catálogo de quasares QDR7 e os $z_{photo}$ calculados pelo ANNz utilizando apenas quatro filtros ( <i>g</i> , <i>r</i> , <i>i</i> , e <i>z</i> ) SDSS. A escala de cor mostra a quantidade de objetos em cada ponto. . . . .	49
3.4	Histograma da frequência de melhor ajuste de cada SED aos quasares QDR7. . . . .	52
3.5	Distribuição das SEDs melhor ajustadas aos quasares do QDR7 em função do <i>redshift</i> espectroscópico do catálogo. A escala de cor mostra a quantidade de objetos em cada ponto. . . . .	52

3.6	Comparação entre os $z_{spec}$ do catálogo de quasares QDR7 e os $z_{photo}$ calculados pelo LePhare utilizando todos os cinco filtros SDSS. A escala de cor mostra a quantidade de objetos em cada ponto. . . . .	53
3.7	Comparação entre os $z_{spec}$ do catálogo de quasares QDR7 e os $z_{photo}$ calculados pelo LePhare utilizando apenas quatro filtros ( $g, r, i, e z$ ) SDSS. A escala de cor mostra a quantidade de objetos em cada ponto.	54
4.1	Subconjunto de SEDs de estrelas presentes no LePhare. Estão representadas 8 SEDs do conjunto de Pickles (1998). . . . .	58
4.2	Conjunto de SEDs de galáxias presentes no LePhare. Todas as SEDs da biblioteca estão representadas. . . . .	59
4.3	Histogramas da frequência do ajuste das SEDs de estrela aos quasares do QDR7 classificados erroneamente como estrelas pelo LePhare. No painel da esquerda está o resultado obtido na presença da banda $u$ . No painel da direita está o resultado sem a banda $u$ . . . . .	60
4.4	Histogramas da frequência do ajuste das SEDs de galáxia aos quasares do QDR7 classificados erroneamente como galáxias pelo LePhare. No painel da esquerda está o resultado obtido na presença da banda $u$ . No painel da direita está o resultado sem a banda $u$ . . . . .	61
4.5	Histogramas da frequência do ajuste das SEDs de quasar aos objetos do QDR7 classificados corretamente como QSOs pelo LePhare. No painel da esquerda está o resultado obtido na presença da banda $u$ . No painel da direita está o resultado sem a banda $u$ . . . . .	61
5.1	Diagramas cor-cor de $g - r \times r - i$ para os objetos dos catálogos DC6B_S_66 de estrelas, DC6B_G_66 de galáxias e DES_66 de quasares. Nos painéis superiores, as cores foram obtidas com base nas magnitudes das tabelas verdade sem acréscimo de erros fotométricos. Nos painéis inferiores as cores são calculadas com base em magnitudes com erros fotométricos. . . . .	66
5.2	O mesmo que na figura 5.1 para $r - i \times i - z$ . . . . .	67
5.3	O mesmo que na figura 5.1 para $i - z \times z - Y$ . . . . .	68
5.4	Comparação entre $z_{photo}$ e $z_{spec}$ . Resultado obtido com o código ANNz para o catálogo DES_66. . . . .	72
5.5	Comparação entre $z_{photo}$ e $z_{spec}$ . Resultado obtido com o código ANNz para o catálogo DES_66 com amostras de treinamento e validação compostas por 1% da amostra alvo. . . . .	73
5.6	Comparação entre $z_{photo}$ e $z_{spec}$ . Resultado obtido com o código LePhare para o catálogo DES_66. . . . .	74

B.1	Distribuição de $\Delta m_g$ em função da magnitude simulada na banda $g$ . No painel da esquerda vemos em azul a distribuição de $\Delta m_g$ para as estrelas do DC6B e, sobrepostas em vermelho, as médias com barras de $1\sigma$ calculadas após a remoção dos objetos que apresentavam $\Delta m_g$ fora do limite de $3\sigma$ . No painel da direita está a distribuição de $\Delta m_g$ simulada para quasares em uma área de 5000 graus quadrados no filtro $g$ do DES. A barra colorida representa o logartimo da contagem de objetos em cada ponto. . . . .	84
B.2	O mesmo que na figura B.1 para a banda $r$ . . . . .	85
B.3	O mesmo que na figura B.1 para a banda $z$ . . . . .	85
B.4	O mesmo que na figura B.1 para a banda $Y$ . . . . .	85
B.5	Erros fotométricos na banda $g$ em função da magnitude observada na mesma banda. No painel da esquerda vemos em azul a distribuição de erro para as estrelas do DC6B e, sobrepostas em vermelho, as medianas e as barras de $1\sigma$ . No painel da direita está a distribuição de erros fotométricos simulada para quasares em uma área de 5000 graus quadrados no filtro $g$ do DES. A barra colorida representa o logartimo da contagem de objetos em cada ponto. . . . .	86
B.6	O mesmo que na figura B.5 para a banda $r$ . . . . .	86
B.7	O mesmo que na figura B.5 para a banda $z$ . . . . .	87
B.8	O mesmo que na figura B.5 para a banda $Y$ . . . . .	87

# Lista de Tabelas

1.1	Densidades numéricas de galáxias ativas e não-ativas por volume no Universo próximo (tabela reproduzida de Mo, van den Bosch e White 2010). . . . .	5
2.1	Constantes e valores de melhor ajuste para a QLF de Richards et al. (2006) na forma de uma lei de potência. . . . .	16
2.2	Constantes e valores de melhor ajuste para a QLF de Willott et al (2010). . . . .	18
2.3	Trecho do catálogo de saída do módulo que integra as QLFs. . . . .	22
2.4	Trecho do catálogo de saída do módulo que gera as magnitudes aparentes. No caso, as magnitudes foram geradas nos filtros SDSS. . . . .	29
2.5	Características do catálogo SDSS_8417. . . . .	32
2.6	Características o catálogo DES_5000. . . . .	39
2.7	Número de objetos por banda por intervalo de <i>redshift</i> no catálogo DES_5000 limitado em magnitude. . . . .	41
3.1	Métricas do resultado do cálculo de <i>photo-z</i> do ANNz para o catálogo QDR7 com cinco e quatro bandas. . . . .	50
3.2	Métricas do resultado do cálculo de <i>photo-z</i> do LePhare para o catálogo QDR7 com cinco e quatro bandas. . . . .	53
4.1	Porcentagem de classificação dos quasares do catálogo QDR7 segundo o menor $\chi^2$ obtido com o LePhare. . . . .	59
4.2	Identificação numérica dos tipos de SEDs de estrelas que podem ser ajustadas pelo algoritmo LePhare para classificar objetos. . . . .	60
4.3	Identificação numérica dos tipos de SEDs de galáxias que podem ser ajustadas pelo algoritmo LePhare para classificar objetos. . . . .	60
5.1	Características gerais da geração dos catálogos simulados utilizados nos testes das ferramentas de análise e dos catálogos de quasares da literatura (QDR7 e QR09). . . . .	64
5.2	Contagem de <i>dropouts</i> por banda presentes no catálogo DES_5000. . . . .	65



5.3	Contagem de quasares, estrelas e galáxias <i>dropouts</i> nas bandas <i>i</i> e <i>z</i> . . .	69
5.4	Classificações possíveis para cada tipo de objeto. . . . .	70
5.5	Classificação dos quasares do catálogo DES_66 segundo a menor $\chi^2$ obtida com o LePhare. . . . .	70
5.6	Classificação das estrelas e galáxias dos catálogos DC6B_S e DC6B_G segundo o menor $\chi^2$ obtido com o LePhare. . . . .	71
5.7	Métricas do resultado do cálculo de photo-z do ANNz para o catálogo simulado pelo AddQSO com diferentes tamanhos de amostra de treinamento. . . . .	73
5.8	Métricas do resultado do cálculo de photo-z do LePhare para o catálogo DES_66. . . . .	74

# Capítulo 1

## Introdução

Até meados do século XX, galáxias eram consideradas objetos cuja emissão era dominada pelas estrelas. Sendo assim, supondo que as atmosferas estelares estão em equilíbrio térmico e hidrodinâmico, a emissão de uma galáxia é de origem térmica e pode ser aproximada por uma soma de espectros de Planck de diferentes temperaturas equivalentes àquelas das estrelas que a compõe. Uma vez que o intervalo de temperatura apresentado por estrelas de diferentes massas e em diversas etapas evolutivas não se estende muito além de  $3000\text{ K} < T < 40000\text{ K}$ , o espectro de uma galáxia de emissão predominantemente estelar irá estar relativamente confinado a  $4000\text{ \AA} < \lambda < 20000\text{ \AA}$  (galáxias que apresentam surto de formação estelar e poeira estendem a faixa de comprimentos de onda em direção ao ultravioleta e ao infravermelho distante, respectivamente).

O paradigma começou a mudar em 1943, quando uma análise sistemática feita por Carl Seyfert chamou atenção para uma nova classe de objetos, hoje conhecidos como "galáxias Seyfert" que apresentam brilho superficial muito alto em seus núcleos. Além disso, suas regiões centrais são dominadas por linhas de emissão de grande excitação e algumas de grande largura. De fato, a energia de excitação necessária para criar essas linhas denotam que os átomos emissores foram excitados por fótons mais energéticos que aqueles produzidos por estrelas jovens responsáveis, por exemplo, pela ionização de regiões H II.

Na década de 60, foi feito um levantamento em rádio denominado *Third Cambridge Catalogue of Radio Sources* (3C - Edge et al. 1959), com observações na faixa de 158 MHz. Posteriormente, ele foi ampliado levando em conta novas observações em 178 MHz, dando origem ao catálogo de radiofontes *Third Cambridge*

*Catalogue of Radio Sources Revised* (3CR - Bennett 1962). Muitas das fontes desses catálogos foram atribuídas a galáxias próximas, porém a baixa resolução espacial dos radiotelescópios nessas frequências resultava numa grande incerteza na posição, o que tornava a identificação de contrapartidas ópticas muito difícil. Na verdade, a identificação óptica completa do catálogo 3CR só foi finalizada nos anos 90, com o aumento em resolução angular devido a técnicas de interferometria. Apesar das dificuldades, em 1963, Thomas Matthews e Allan Sandage conseguiram relacionar o objeto 3C48 a uma fonte pontual com magnitude  $V \cong 16$  (Matthews & Sandage 1963)<sup>1</sup>. O espectro dessa fonte consiste de um contínuo azul e linhas de emissão fortes e largas que, na época, não foram relacionadas a transições atômicas na faixa do óptico. No mesmo ano, Maarten Schmidt fez uso de uma ocultação da fonte 3C273 pela Lua para determinar sua posição exata e, conseqüentemente, a relacionou com uma fonte pontual no óptico que também apresentava espectro de características nunca antes vistas naqueles comprimentos de onda (Schmidt, 1963). Schmidt identificou as fortes linhas de emissão como a série de Balmer, o que significava que a fonte deveria apresentar um *redshift*,  $z$ , de  $z \cong 0,16$ . Considerando a validade da lei de Hubble<sup>2</sup> e interpretando o desvio para o vermelho como um *redshift* cosmológico, 3C273 deve estar localizado a distância de  $D \approx 500h^{-1}$  Mpc, onde  $h$  é um número sem dimensão que parametriza a constante de Hubble<sup>3</sup>. Essa distância implica numa magnitude absoluta de  $M_B = -25,3 + 5 \log h$  que é cerca de  $\approx 100$  vezes mais brilhante que uma galáxia espiral típica.

Após este resultado e com a melhora na determinação da posição de fontes rádio, outros espectros que permaneciam com linhas não identificadas foram associados a objetos com grande *redshift*. De aspecto estelar e apresentando variabilidade na faixa óptica e no rádio, esses objetos foram inicialmente chamados de "superestrelas". Posteriormente, eles foram renomeados de *quasi-stellar radio sources* e, em 1964, o termo "quasares" foi criado por Hong-Yee Chiu em um artigo publicado na *Physics Today* (Hong-Yee Chiu, 1964).

<sup>1</sup>O objeto 3C48 teve seu *redshift* posteriormente calculado como  $z = 0,3675$  por Greenstein & Matthews (1963).

<sup>2</sup>A lei de Hubble é uma relação entre a velocidade de recessão ( $v$ ) de um corpo fora da Via Láctea e sua distância ( $d$ ):  $v = H \times d$ , onde  $H$  é a chamada constante de Hubble. Segundo medidas com base na anisotropia da radiação cósmica de fundo utilizando o WMAP, o valor de  $H$  é de  $70,2 \pm 1,4 \text{ kms}^{-1}\text{Mpc}^{-1}$  (Komatsu et al. 2010).

<sup>3</sup> $H = 100h \text{ kms}^{-1}\text{Mpc}^{-1}$

A grande emissão em ultravioleta dos quasares, e, por consequência, a sua cor muito azulada levou à realização de levantamentos ópticos a procura de objetos puntiformes com excesso de radiação UV. Um exemplo é a amostra do *Bright Quasar Survey*, parte do levantamento Palomar-Green (Schmidt & Green 1983), que é definida pelos limites de magnitude ( $B < 16,2$ ) e cor ( $U - B < -0,4$ ). Essa seleção por cor levou à descoberta de fontes que apresentam uma distribuição de energia com excesso na parte azul do espectro, além de linhas de emissão fortes e largas, e grandes *redshifts*, o que as tornava muito similares aos quasares exceto por uma propriedade: não apresentavam emissão em rádio. Esses objetos, que são mais numerosos que os quasares por um fator de pelo menos 10, passaram a ser conhecidos como quasares "*radio-quiet*", radioquiescentes (enquanto que os quasares que tinham contrapartidas em rádio receberam a nomenclatura "*radio-loud*", radiosonoros) ou, simplesmente, QSOs (quasi-stellar objects). Hoje em dia, porém, essa terminologia foi abandonada, uma vez que até mesmo QSOs apresentam emissão em rádio quando observados com sensibilidade suficientemente alta. Sendo assim, a separação *radio-loud/radio-quiet* não é mais considerada e o que se encontra na literatura é a utilização indistinta de ambos os termos: quasares ou QSOs.

Apesar dos quasares serem fontes com aspecto majoritariamente pontuais, é possível, em alguns casos, observar a presença dos jatos relativísticos. Eles são os objetos mais luminosos do Universo e muitos já foram encontrados a grandes distâncias<sup>4</sup>. Schneider (2006) e Sparke & Gallagher (2007) mencionam que quase todos os quasares excedem suas galáxias hospedeiras em brilho e apenas nos mais próximos e fracos podemos resolvê-las. Na maior parte dos casos em que o imageamento da galáxia hospedeira dos quasares foi possível, não se obteve detalhes para caracterizar as morfologias. Apesar disso, parece que as hospedeiras de quasares de baixo *redshift* se mostraram preferencialmente elípticas ou sistemas fortemente perturbados e em interações (e.g., Canalizo et al. 2007).

Os quasares, bem como as galáxias Seyfert, fazem parte de um grupo de objetos conhecidos como "galáxias com núcleo ativo". Essas galáxias mostram uma distribuição de energia cuja intensidade se mantém praticamente constante por quase todos os comprimentos de onda, o que não é observado em galáxias normais. De uma

---

<sup>4</sup>O quasar mais distante encontrado até a presente data possui  $z = 7,085$  (Mortlock et al. 2011).

pequena região central dessas galáxias, denominada "núcleo ativo de galáxia" (*active galactic nucleus* - AGN), é observada uma emissão que pode englobar desde comprimentos de onda rádio até os de raios-X e até mesmo o regime dos raios gama, o que sugere uma fonte de radiação não térmica. Na figura 1.1, de Schneider (2006), vemos uma comparação entre os espectros do quasar 3C273 e o de uma galáxia elíptica não-ativa, onde fica claro que a emissão do quasar é observada em todo o intervalo da radiação eletromagnética, enquanto que a radiação da galáxia elíptica tem sua radiação concentrada numa pequena faixa. O espectro de um QSO é basicamente composto por um contínuo, que pode ser representado por uma lei de potência, com fortes linhas de emissão e absorção.

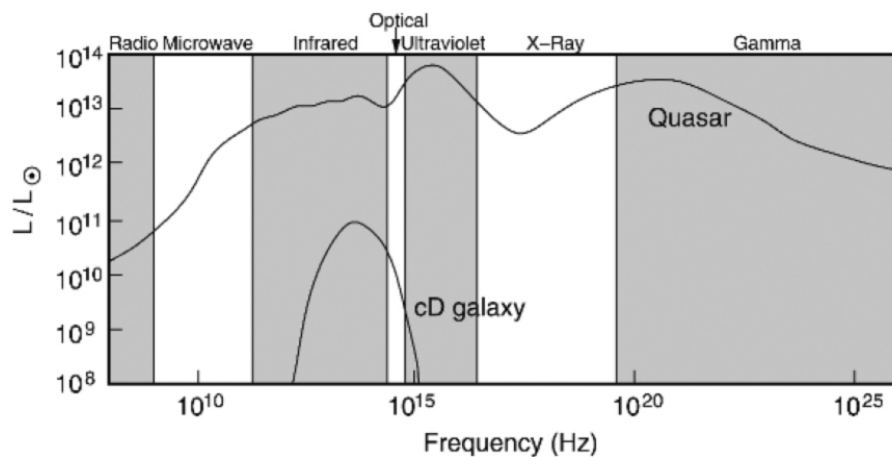


Figura 1.1: Espectro do quasar 3C273 em comparação com o espectro de uma galáxia elíptica não-ativa de tipo morfológico cD. A emissão da radiação do QSO é observada em quase toda a extensão do espectro eletromagnético com a energia por intervalo de frequência logarítmica sendo quase constante (Schneider, 2006).

Acredita-se que galáxias ativas compõem uma família de diferentes tipos de AGN que diferem em suas propriedades espectrais, suas luminosidades e a razão entre a luminosidade do núcleo e da luz gerada pelas estrelas que compõe a galáxia. Elas podem ser separadas em quatro grandes classes, que podem ser afetadas pelo comprimento de onda de observação da fonte: galáxias Seyfert<sup>5</sup>, radiogaláxias, blazares e quasares, cujas densidades numéricas de ocorrência no Universo próximo bem como uma comparação com as densidades de galáxias não-ativas podem ser

<sup>5</sup>As galáxias Seyfert podem ser divididas em dois sub-tipos, 1 e 2. O critério de separação é a largura das linhas de emissão: Seyferts 2 apresentam apenas linhas estreitas ( $\text{FWHM} < 1000 \text{ km s}^{-1}$ ) (e.g., Antonucci 1993). O mesmo critério pode ser aplicado às radiogaláxias, dando origem às sub-classes de linhas largas, *broad-line radio galaxies* (BLRG), e de linhas estreitas, *narrow-line radio galaxies* (NLRG).

vistas na tabela 1.1.

Tabela 1.1: Densidades numéricas de galáxias ativas e não-ativas por volume no Universo próximo (tabela reproduzida de Mo, van den Bosch e White 2010).

Tipo de objetos	Densidade numérica [Objetos Mpc <sup>-3</sup> ]
Galáxias de campo	10 <sup>-1</sup>
Espirais luminosas	10 <sup>-2</sup>
Galáxias Seyfert	10 <sup>-4</sup>
Radiogaláxias	10 <sup>-6</sup>
Quasares radioquiescentes	10 <sup>-7</sup>
Quasares radiosonoros	10 <sup>-9</sup>

Há várias evidências nas diferentes classes de AGN que indicam que a grande quantidade de energia sendo gerada é proveniente de uma região pequena. O primeiro exemplo é a presença de largas linhas de emissão, presentes em quasares e galáxias Seyfert 1, que apresentam larguras a meia altura (*full width at half maximum* - FWHM) da ordem de 10000 km s<sup>-1</sup> (e.g., Antonucci 1993), o que, indiretamente, é indicativo de que a luminosidade é proveniente de uma região com raio inferior a 1 pc. Outro exemplo é a presença de variabilidade nas fontes. No caso de galáxias Seyfert, não é incomum encontrar variabilidade na radiação proveniente do núcleo dessas galáxias que em menos de um ano apresentam variação superior a um fator dois, o que é um indicativo de que a área emissora de radiação tem diâmetro menor que 0,3 pc (e.g., Mo, van den Bosch e White 2010). Já os OVV's (*optically violent variables*), uma sub-classe dos Blazares, apresentam variabilidade violenta inclusive na faixa de radiação gama em uma escala de tempo típica que pode chegar a apenas dias (e.g., Mo, van den Bosch e White 2010).

Hoje, acredita-se que as diversas classes de AGN são intrinsecamente similares, sendo as diferenças explicadas por efeitos de geometria ou de projeção decorrentes da observação de componentes distintos do núcleo ativo. Isso constitui o que hoje é conhecido como modelo unificado de AGNs. Nesse modelo, o paradigma mais aceito para explicar a grande quantidade de energia proveniente de uma região tão pequena é a existência de um buraco negro super-massivo (*supermassive black hole* - SMBH) no centro das galáxias ativas. A idéia de SMBHs foi originalmente proposta por Salpeter (1964), Zel'dovich & Novikov (1964) e Lynden-Bell (1969) e, na literatura atual, a idéia mais aceita é que toda galáxia contém um SMBH em seu centro (Kormendy & Richstone, 1995). A diferença entre uma galáxia ativa ou inativa

reside apenas no abastecimento do SMBH.

Além do SMBH, o modelo unificado inclui os seguintes componentes:

- um disco de acreção de gás e poeira;
- uma região de linhas largas circundante ao disco de acreção com tamanho típico de 0,3 pc (Mo, van den Bosch e White 2010), densidade de elétrons da ordem de  $n_e \approx 3 \times 10^{-9} \text{ cm}^{-3}$  e temperatura em torno de  $T \approx 20000 \text{ K}$  (Schneider et al. 2006). A largura das linhas é decorrente de alargamento Doppler produzido pela rotação kepleriana do gás;
- uma região de linhas estreitas composta por gás de densidade inferior a  $10^6 \text{ cm}^{-3}$  localizada ao redor de  $\approx 50 \text{ pc}$  do núcleo e que pode ser resolvida espacialmente em galáxias ativas próximas (observações com o Hubble Space Telescope mostraram, em algumas Seyferts 2, um meio de gás turbulento - Capetti et al., 1996; Bennert et al., 2002);
- um torus de matéria interior à região de linhas estreitas composto por uma distribuição de poeira que o torna geometricamente espesso (técnicas de interferometria são capazes de imagear as regiões centrais de AGNs e confirmam a presença de uma estrutura toroidal composta por poeira - Tristram et al. 2007; Raban et al., 2009);
- jatos relativísticos que são fluxos altamente colimados de matéria.

Dentro desse cenário, AGNs obscurecidos de tipo 2 como os presentes nas Seyferts 2 se mostram como tal pois o observador tem o SMBH escondido pelo torus de poeira, enquanto que os AGNs de tipo 1 como os SMBHs das Seyferts 1 são observados diretamente, sem interferência do torus. A figura 1.2 mostra um esquema do modelo unificado de AGN e as linhas de visada através das quais os diferentes tipos de AGN são observados.

A energia dos AGNs é provida pela energia potencial cedida pelo gás quando ele cai no SMBH. Schneider (2006) explica que pelo princípio de acreção, um gás caindo em um objeto compacto como um SMBH cede energia potencial primeiramente na forma de energia cinética. Contudo, devido à conservação de momento angular do gás, este não pode cair diretamente no SMBH. A transferência de momento entre

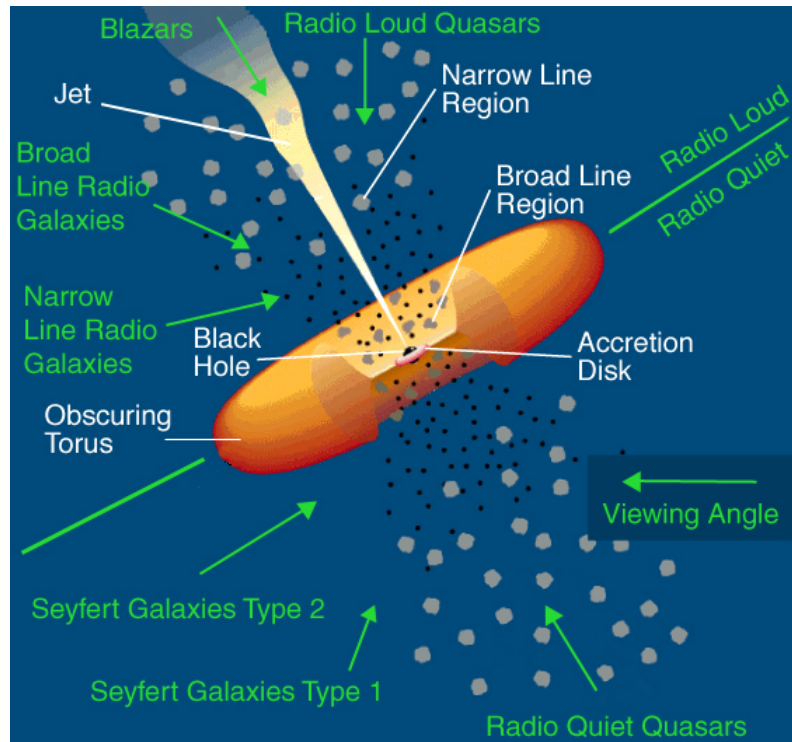


Figura 1.2: O modelo unificado de núcleos ativos de galáxias. Ao redor do buraco negro supermassivo está orbitando um disco de acreção de gás e poeira. Frequentemente, há a presença de jatos de partículas alinhados com o eixo de rotação do buraco negro. O tipo de AGN a ser observado da Terra depende do ângulo de visada (crédito da imagem: NASA apud Pierre Auger Observatory - [http://www.auger.org/news/PRagn/about\\_AGN.html](http://www.auger.org/news/PRagn/about_AGN.html)).

as partículas do gás fará com que esse assuma a forma de um disco perpendicular à direção do vetor momento angular. Esse disco será aquecido por fricção interna devido à rotação diferencial que também ocasionará uma pequena desaceleração da velocidade de rotação, fazendo com que o gás se mova para dentro, transformando a energia potencial em cinética. Uma vez que a força gravitacional exercida pelo SMBH é muito superior às forças de fricção no gás, o disco irá rotacionar seguindo aproximadamente um movimento kepleriano. De acordo com o teorema do virial, metade da energia potencial liberada é convertida em energia cinética, enquanto que a outra metade pode ser convertida em energia interna.

A compreensão das propriedades e do processo de formação de galáxias com núcleo ativo é de grande relevância para o estudo da formação de galáxias em geral. A existência da correlação entre as massas do SMBH e as de suas galáxias hospedeiras (Magorrian et al. 1998) sugere que a formação dos SMBHs está intimamente ligada à formação de galáxias (e.g., Lemastra et al. 2010). Já o fato de,



conforme mencionado, todas as galáxias conterem um SMBH em seus centros, nos leva a questionar se todas as galáxias foram ativas em alguma fase de sua evolução e, caso negativo, o que leva à atividade em apenas alguns casos. Ainda, sendo grandes fontes de energia, acredita-se que AGNs tenham grande impacto na evolução de suas galáxias hospedeiras. Esse *feedback* de energia deve ser levado em conta em qualquer teoria de formação e evolução de galáxias. Em grandes *redshifts*, o estudo de AGNs é chave para a compreensão da formação de estrutura no Universo jovem. Entre as questões que podem ser estudadas podemos citar: a contribuição dos AGNs à reionização, ao aquecimento do meio intergaláctico e os efeitos sobre formação de estruturas (e.g., Boutsia et al. 2011; Mitra et al. 2011); cenários da formação buracos negros sementes que virão a crescer e se transformar no SMBHs presentes nos centros de quase todas as galáxias (e.g., Volonteri & Rees, 2005); a autoregulação e cessação da atividade de formação estelar em galáxias massudas (e.g., Bundy et al. 2008).

Devido à grande luminosidade apresentada pelos quasares e a consequente possibilidade de observá-los a enormes distâncias, esses objetos compõe a classe mais interessante e de maior utilidade entre as galáxias de núcleo ativo a ser estudada para compreendermos o Universo. A problemática do processo de formação dos SMBHs ganha uma outra dimensão quando é adicionada a existência de quasares em *redshifts* superiores a 7, o que torna necessário um processo de formação que seja rápido o suficiente para produzir e abastecer um SMBH quando o Universo tinha apenas algumas centenas de milhões de anos. A descoberta dos quasares em  $z > 6$  mostrou um Universo altamente ionizado já nessa época (Becker et al. 2007), sem evidências de reionização. O mecanismo de abastecimento dos SMBHs também pode ser estudado através dos quasares, onde são necessárias várias massas solares ( $M_{\odot}$ ) por ano de combustível sendo fornecido por  $10^7$  anos ou mais (e.g., Martini 2004).

Um SMBH em um quasar luminoso é capaz de crescer, durante seu período de maior atividade, uma massa da ordem de  $10^8 M_{\odot}$ . Por esse motivo, fusões de galáxias ricas em gás parecem ser o cenário mais viável de abastecimento. Apesar de haver alguns exemplos de quasares em sistemas em processo de fusão (e.g., Bahcall et al. 1997) e de esse cenário ser corroborado pelos modelos de coevolução de buracos negros e galáxias (e.g., Hopkins et al. 2006), a emissão intensa dos quasares

torna difícil a classificação da morfologia da galáxia hospedeira, o que dificulta a confirmação desse modelo de fusão.

Uma forma de estudarmos a evolução dos quasares é através da determinação da função de luminosidade dos QSOs. A QLF é definida como a densidade numérica de quasares por intervalo de luminosidade (ou de magnitude absoluta). Quando observada por levantamentos profundos, a QLF apresenta uma quebra em uma dada luminosidade característica, que torna a curva mais plana para objetos mais fracos. Sendo assim, a forma mais frequente de representar a QLF é através de uma lei de dupla potência (Boyle et al. 2000; Croom et al. 2004; Richards et al. 2006) que, em função da magnitude, é equivalente a:

$$\Phi(M, z) = \frac{\Phi(M^*)}{10^{0,4(\alpha+1)[M-M^*(z)]} + 10^{0,4(\beta+1)[M-M^*(z)]}}, \quad (1.1)$$

onde  $\alpha$  descreve a inclinação da curva que representa a parte fraca da QLF e  $\beta$  a inclinação da curva que representa a parte brilhante. O parâmetro  $\Phi(M^*)$  tem unidade de densidade numérica e é responsável pela normalização da QLF. A observação em magnitudes inferiores à  $M^*$  permite o estudo do tipo de evolução sofrida pelos quasares. Se a luminosidade dos objetos varia com o tempo, mas sua densidade numérica permanece constante, a evolução é em luminosidade (*pure luminosity evolution* - PLE). Por outro lado, se a densidade numérica dos quasares variar, mas a luminosidade dos objetos permanecer inalterada, a evolução é em densidade (*pure density evolution* - PDE). Trabalhos realizados com levantamentos não muito profundos como os de Boyle et al (2000) e Croom et al (2004) encontraram bons ajustes a modelos de PLE. Porém, esses levantamentos eram limitados a baixos regimes de *redshift*. O que se acredita é que uma parametrização que leva em conta apenas uma evolução em luminosidade não é suficiente para ajustar uma QLF até *redshifts* maiores (e.g., Richards et al. 2006). Há uma clara diminuição na densidade a partir de  $z > 2,5$  (Osmer 1982; Warren et al. 1994; Schmidt et al. 1995; Fan et al. 2001; Richards et al. 2006; Croom et al. 2009), que faz com que o intervalo de *redshifts* entre 2 – 3 seja conhecido como a "época dos quasares" e que impede a aplicação de uma parametrização PLE. Uma forma híbrida entre PLE e PDE foi utilizada em levantamentos em raios-X para descrever a QLF nessa faixa do espectro (Schmidt & Green 1983; Ueda et al. 2003).

A evolução e a forma da QLF com  $z \geq 6$  nos dá informações sobre o crescimento de SMBHs e o papel dos quasares na reionização. O *Sloan Digital Sky Survey* (SDSS - York et al. 2000) foi capaz de realizar excelentes medidas na evolução da faixa brilhante da QLF, mostrando que esta pode ser ajustada por uma lei de potência até o limite de  $M_{1450} \leq -23,3$  (Fan et al. 2001; Richards et al. 2006). Ao mesmo tempo, levantamentos feitos em uma área menor porém com maior profundidade sugerem um achatamento da função no regime de baixa luminosidade (Shankar & Mathur 2007; Willot et al. 2009). O levantamento que mais contribui para essa parte menos luminosa da função de luminosidade foi o do observatório espacial Chandra (Weisskopf et al. 2000) com seus *Deep Fields* (e.g., Giacconi et al. 2002; Hornschemeier et al. 2000), que encontraram resultados consistentes com uma proporcionalidade entre a massa do SMBH e do halo de matéria escura. Estimativas baseadas nesse resultado sugerem que a contribuição dos quasares para a reionização é da ordem de 10% em  $z \approx 6$  e da ordem de 30% quando integrada entre  $5,5 < z < 6,5$  (Shankar & Mathur 2007). Com levantamentos mais profundos é possível vincular melhor a parte fraca da função de luminosidade dos quasares e a contribuição desses objetos para a reionização.

De acordo com Mortlock et al. (2008) e Venemans et al. (2007), o fato dos quasares serem não-transientes e a possibilidade de seus espectros poderem ser obtidos com alta razão sinal/ruído confere a oportunidade de avaliação do material que preenche o espaço percorrido pela luz emitida pelo quasar. Com isso, podemos determinar o conteúdo de matéria bariônica e leptônica, a metalicidade, a temperatura, o grau de ionização, entre outras propriedades físicas. Desde as primeiras observações espectroscópicas de quasares em  $z \approx 2$ , ficou claro que o gás que compõe o meio intergaláctico é altamente ionizado (Gunn & Peterson, 1965) e que o conteúdo de matéria bariônica do Universo atual é dominado por essa matéria ionizada e não pelas estrelas (e.g., Fukugita, Hogan & Peebles, 1998). Com a obtenção de espectros de quasares cada vez a mais altos *redshifts*, fica claro que o meio intergaláctico é ionizado já nessas épocas. Resultados de polarização obtidos com o WMAP apontam um *redshift* de  $z \approx 8,2$  para a época de reionização (Dunkley et al. 2009). Por esse motivo, é de extrema importância que se aumente o tamanho da amostra de quasares de  $z > 6,5$  para que possamos avaliar épocas anteriores àquelas que

a transmissão na floresta de Ly $\alpha$  já indica ionização. Isso será possível através de levantamentos profundos que incluam bandas no infravermelho.

O *Dark Energy Survey* (DES) é um dos exemplos de levantamentos que irão permitir estudos desse tipo. O DES é um levantamento fotométrico que se estenderá do óptico ao infravermelho próximo utilizando as bandas  $g$ ,  $r$ ,  $i$ ,  $z$  e  $Y$ . A proposta do levantamento é de, através do uso de quatro experimentos distintos e independentes, sendo eles a abundância e distribuição espacial de aglomerados de galáxias, o lenteamento fraco, as oscilações acústicas de bárions e as supernovas tipo Ia (Annis et al. 2005), impor vínculos à equação de estado da energia escura a 6% e sua evolução a 20%.

O DES cobrirá 5000 graus quadrados da calota galáctica Sul (Abbott et al. 2005) e em adição aos filtros mencionados o levantamento *VISTA Hemisphere Survey* (VHS) irá prover dados nas bandas  $J$ ,  $H$  e  $K$ . Além disso, aproximadamente 4000 graus quadrados da região observada irão se sobrepor a região do levantamento *South Pole Telescope* (Ruhl et al. 2004).

O equipamento usado é uma nova câmera com 62 CCDs, a DECam, montada no telescópio Blanco de 4 metros localizado no *Cerro Tololo Inter-American Observatory* - CTIO. De acordo com o planejamento, o levantamento terá 525 noites de observações distribuídas por cinco anos nos períodos de setembro a fevereiro (Abbott et al. 2005). Os limites de fotometria esperados são:  $g = 25,21$ ,  $r = 24,77$ ,  $i = 24,04$ ,  $z = 23,44$  e  $Y = 21,70$ , uma vez completado o levantamento. A primeira luz ocorreu no início de setembro de 2012 e a fase de comissionamento já se encontra encerrada. A fase de verificação científica terminou em 22 de fevereiro de 2013 e produziu dados em mais de 150 graus quadrados atingindo profundidades próximas à máxima esperada pelo levantamento. Apesar de ser voltado para o estudo da energia escura, a grande profundidade e área cobertas pelo DES irão proporcionar uma imensa amostra estatística, aumentando o tamanho da amostra de quasares com  $z > 4$  e identificando uma quantidade significativa de objetos com  $z > 6$ . Com sua amostra de quasares, o DES irá proporcionar a possibilidade de se estudar a demografia desses objetos como suas propriedades de aglomeração, as condições do meio intergaláctico próximo à época de reionização, a existência de SMBHs no Universo jovem, a coevolução entre SMBHs e os halos de matéria escura, entre outros.

A proposta desse trabalho é simular um catálogo de quasares capaz de prever a quantidade desses objetos e suas distribuições de magnitudes e de *redshift*, no DES. Além de fazer previsões para o levantamento, nosso objetivo é testar ferramentas de análise, de identificação e de cálculo de *redshift* fotométrico (*photo-z*) de quasares e examinarmos suas eficiências. No capítulo 2 descrevemos o funcionamento da suíte de algoritmos AddQSO que criamos para gerar as simulações. Nesse capítulo também validamos os códigos e fazemos previsões para o DES. No capítulo 3 descrevemos dois métodos de cálculo de *photo-z*, redes neurais e ajustes de modelos, que testamos utilizando quasares espectroscopicamente confirmados do SDSS. No capítulo 4 discutimos a seleção por cor de quasares em uma amostra que contenha estrelas e galáxias e apresentamos o método de ajuste de modelos como uma forma de classificação. No capítulo 5 descrevemos os resultados obtidos nos testes de cálculo de *photo-z* e de classificação com os quasares simulados. Finalmente, no capítulo 6, descrevemos nossas conclusões e discutimos as perspectivas para a continuidade do trabalho.

# Capítulo 2

## O *pipeline* AddQSO

O AddQSO é uma suíte de algoritmos que criamos para simular catálogos de QSOs em um determinado volume do céu e para um dado conjunto de filtros. Os algoritmos fazem uso de uma ou mais funções de luminosidade, um conjunto de filtros e uma lista de distribuições espectrais de energia (*spectral energy distributions* - SEDs). Para objetos de *redshifts* maiores, são incorporados os efeitos da floresta de Ly $\alpha$ .

Composto por duas fases principais e duas fases eletivas, o AddQSO integra uma QLF para obter valores de *redshift* ( $z$ ) e magnitude absoluta na banda  $i$  ( $M_i$ ), e calcula a magnitude aparente  $i$  ( $m_i$ ) para cada objeto. A partir desses dados, gera as magnitudes aparentes nas outras bandas através da convolução entre filtros e SEDs, aplicando as correções de absorção quando necessárias. Então, é possível distribuir os objetos simulados aleatoriamente no espaço e/ou aplicar erros às magnitudes geradas.

O AddQSO serve-se das facilidades do Portal Científico desenvolvido pelos colaboradores do DES-Brazil. O objetivo do Portal é oferecer um ambiente de desenvolvimento colaborativo para a análise científica, onde é possível manter um registro completo de versionamento de códigos e proveniência dos dados. Nele, códigos são integrados em *workflows* a serem executados, de forma paralela ou não, em *clusters* computacionais. Dentro desse contexto, o AddQSO se tornou um *pipeline* onde cada uma de suas fases representa um módulo. Para integrarmos o AddQSO ao Portal, tornamos suas execuções não-supervisionadas. Na figura 2.1 é exibida a estrutura do *workflow* de nosso *pipeline*. A chave de boca vermelha abre uma aba auxiliar para que o usuário escolha as configurações do módulo. O módulo de distribuição

especial ainda não foi acrescentado à estrutura desenvolvida no Portal.

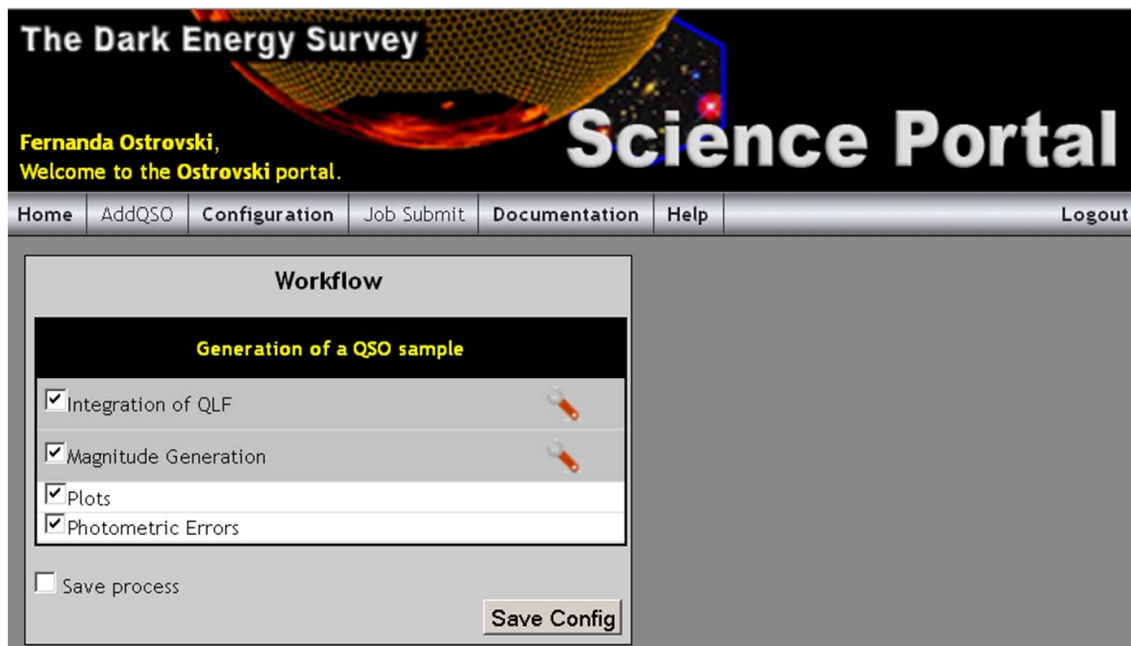


Figura 2.1: Estrutura do *workflow* do *pipeline* AddQSO integrado no Portal Científico do DES.

A integração do AddQSO no Portal permitiu que todos os módulos, sendo algoritmos distintos e escritos em linguagens de programação diferentes (Fortran 77 e Python), pudessem ser combinados de forma que a saída de um dos algoritmos é automaticamente adaptada para se tornar a entrada de um próximo. Além disso, todos os diagramas de análise e tabelas de estatísticas são criados durante a execução do *pipeline*. No final de uma rodada, temos toda a informação de saída necessária, ao mesmo tempo em que mantemos um registro das configurações de entrada e tempo de execução. A ligação entre os diferentes módulos do *workflow* e suas interdependências são definidas através de uma interface criada por um código escrito em xml que é mostrado no apêndice A.1.

Se todas as fases forem realizadas com sucesso, o catálogo final gerado pelo AddQSO irá conter *redshift*, magnitudes aparentes, erros fotométricos, ascensão reta e declinação. O processo de simulação será descrito em detalhes nas seções a seguir, orientadas de acordo com a estrutura do *pipeline*.

## 2.1 Geração das distribuições de $z$ , $M_i$ e $m_i$

O primeiro passo na simulação do catálogo de quasares é a obtenção das distribuições de *redshift* e magnitude absoluta em uma banda. Para tal, é necessário integrar uma QLF. Devido ao grande intervalo de *redshifts* no qual esses objetos se distribuem e na dificuldade em se obter uma amostra completa em toda a faixa, em geral temos QLFs para diferentes intervalos de  $z$ . Em sua versão atual, o AddQSO contém a QLF de Richards et al. (2006) para simular objetos até  $z \approx 5,0$  e a de Willott et al. (2010) para  $z > 5,7$ . O AddQSO permite escolher uma ou ambas QLFs dependendo do tipo de catálogo a ser simulado.

A QLF determinada por Richards et al (2006) se baseia em uma amostra de  $\approx 15000$  quasares dentro de uma área efetiva de 1622 graus quadrados do SDSS *Data Release 3* (Abazajian et al. 2005). A amostra é limitada em  $15,0 < i < 19,1$  para objetos com  $0,3 < z < 3$  e é estendida até  $i = 20,2$  para objetos com  $z > 3$ . A figura 2.2 mostra as distribuições de *redshift* e de  $m_i$  dos objetos da amostra usada para a determinação da QLF. Na distribuição de *redshift* fica evidente os efeitos de seleção de objetos. O algoritmo de seleção de quasares do SDSS apresenta uma diminuição da eficiência em  $z = 2,7$  e  $z = 3,5$  (Schneider et al. 2007), ocasionando os efeitos observados no gráfico da distribuição de *redshifts*.

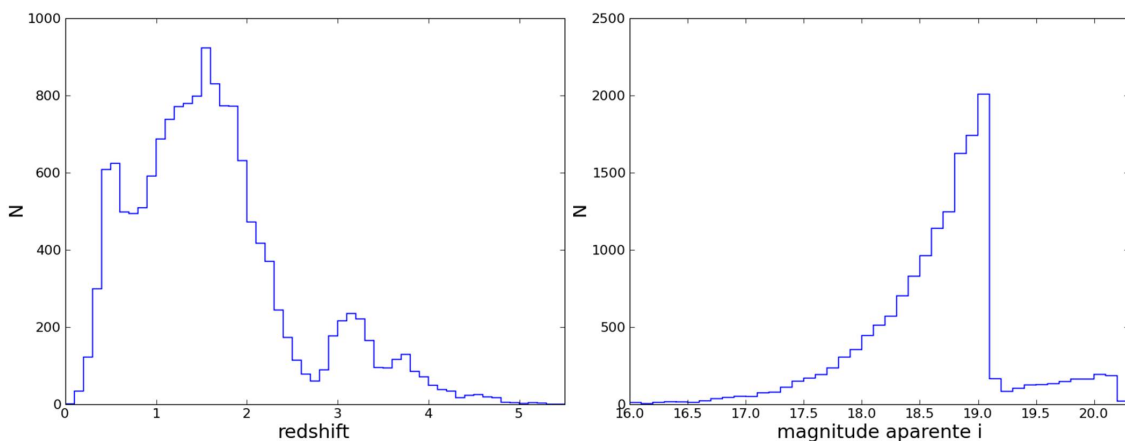


Figura 2.2: Distribuições de *redshift* (painel da esquerda) e de  $m_i$  (painel da direita) para os  $\approx 15000$  quasares da amostra utilizada por Richards et al (2006) na determinação da QLF.

A QLF foi obtida através do método de máxima verossimilhança (MV) ajustada a uma forma parametrizada da função. Como o SDSS não é um levantamento muito profundo, a parametrização escolhida foi de uma lei de potência simples, e os



autores escolheram determinar a MV com respeito a uma forma de PLE, já que essa não difere da PDE devido à falta de uma escala característica em luminosidade. A solução da MV é:

$$\Phi = \Phi^* 10^{\mu [A_1 + A_2(z - z_{ref})]}, \quad (2.1)$$

sendo

$$\mu = M - (M^* + B_1\xi + B_2\xi^2 + B_3\xi^3), \quad (2.2)$$

e

$$\xi = \log \left( \frac{1+z}{1+z_{ref}} \right). \quad (2.3)$$

Os valores de  $\Phi^*$ ,  $A_1$ ,  $A_2$ ,  $B_1$ ,  $B_2$  e  $B_3$  são livres enquanto que  $z_{ref}$  e  $M^*$  são fixados em 2,45 e  $-26$ , respectivamente. Os valores melhor ajustados e utilizados no AddQSO são dados na tabela 2.1. O valor de  $A_2$  varia de acordo com o regime de *redshift* para incluir a variação na inclinação na QLF para objetos de  $z > 2,4$ . A escolha de uma função que varia com  $z$  foi motivada por trabalhos na literatura como os de Schmidt et al. (1995) e Fan et al. (2001), que encontraram um achatamento na inclinação da curva da QLF para grandes *redshifts*. Esse efeito também foi observado por Richards et al. (2006) na determinação de sua QLF. Na figura 2.3 podemos observar a QLF como a linha azul. Nessa figura também estão representadas outras formas de QLF determinadas pelos autores porém não utilizadas pelo AddQSO. Os autores escolheram não utilizar  $z = 0$  como referência para a aplicação das correções-K de seus objetos. Ao invés disso, eles optaram por  $z = 2$ , mais próximo à mediana da distribuição de *redshifts* da amostra. Sendo assim, a magnitude absoluta utilizada por eles está associada à banda  $i$  de uma quasar em  $z = 2$  cujo espectro tem um contínuo de inclinação  $\alpha_\nu = -0,5$  com comprimento de onda de repouso efetivo de  $\approx 2500 \text{ \AA}$ .

Tabela 2.1: Constantes e valores de melhor ajuste para a QLF de Richards et al. (2006) na forma de uma lei de potência.

$z$	$A_1$	$A_2$	$B_1$	$B_2$	$B_3$	$M^{*\ddagger}$	$z_{ref}^{\ddagger}$	$\log(\Phi^*)^\dagger$
$z \leq 2,4$	0,84	0,00	$1,43 \pm 0,04$	$36,63 \pm 0,10$	$34,39 \pm 0,26$	-26	2,45	-5,70
$z > 2,4$	$0,83 \pm 0,01$	$-0,11 \pm 0,01$	$1,43 \pm 0,04$	$36,63 \pm 0,10$	$34,39 \pm 0,26$	-26	2,45	-5,70

$\ddagger$  parâmetro fixo

$\dagger$  QSOs  $\text{Mpc}^{-3} \text{ mag}^{-1}$

A QLF determinada por Willott et al. (2010), por sua vez, é baseada em 40

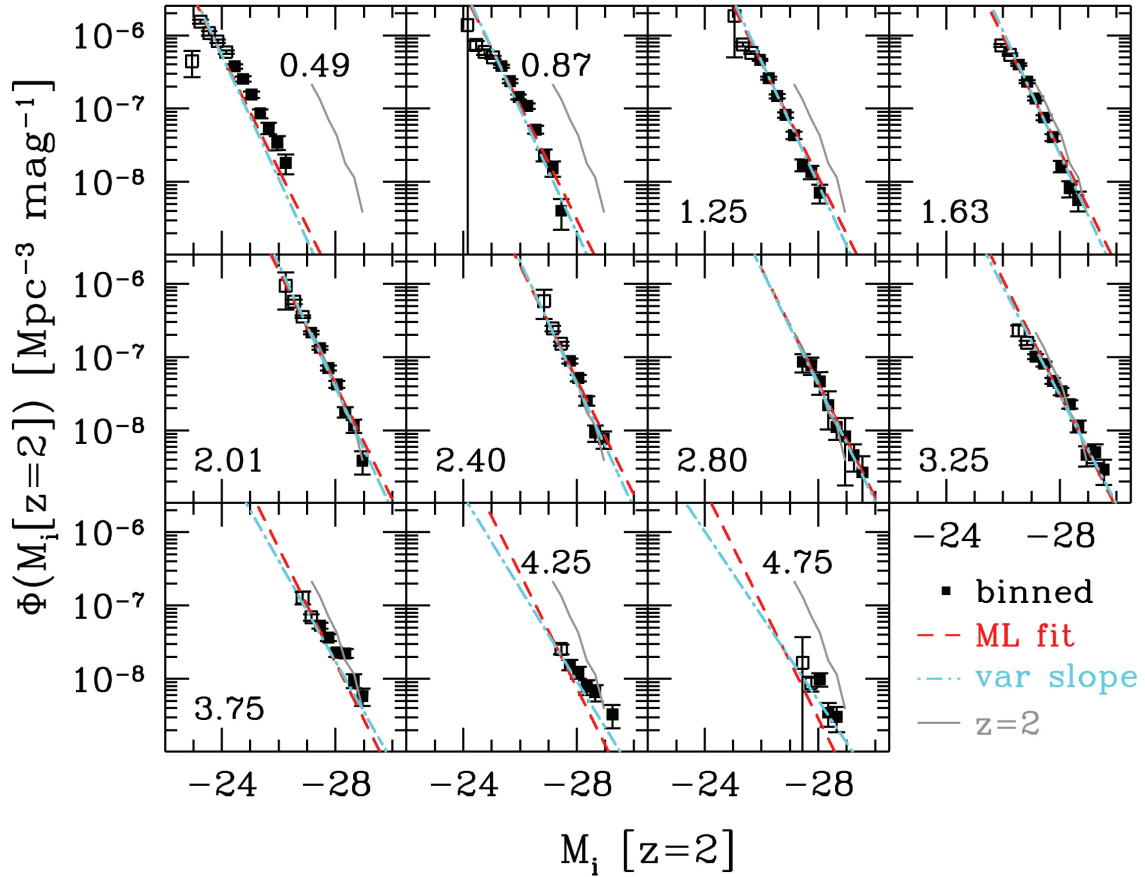


Figura 2.3: QLF de Richards et al. (2006) para diferentes intervalos de *redshift*. A função sendo usada está representada pela linha azul ponto-tracejada. A linha cinza representa a curva a  $z = 2,01$  e foi reproduzida para comparação (Richards et al. 2006). Os números em cada painel mostram o centro dos intervalos de *redshift*, que têm 0,5 de largura.

quasares com *redshifts* entre  $5,74 < z < 6,42$  provenientes dos levantamentos SDSS e *Canada-France High- $z$  Quasar Survey* (CFHQS - Willott et al. 2009). Os quasares proveniente do SDSS vem de duas amostras diferentes. Um total de 14 objetos foi obtido na amostra descrita por Fan et al. (2006) através de um critério uniforme de seleção em uma área total de  $\approx 6600$  graus quadrados do céu. O intervalo de *redshift* é  $5,74 < z < 6,42$  com  $M_z < 20,2$  e cor  $i - z > 2,2$ . Há ainda 10 quasares com *redshifts* entre  $5,78$  e  $6,12$  selecionados em uma pequena porção do SDSS de maior profundidade, onde quasares mais fracos podem ser buscados. Esses objetos são provenientes de uma área de 260 graus quadrados e  $20 < m_z < 21$ . Os quasares restantes fazem parte de uma amostra do CFHQS, cujos critérios de seleção incluem um limite de magnitude  $m_z < 22,5$  e cor  $i - z > 2$ .

A QLF de Willott et al. (2010) utilizada pelo AddQSO é o modelo paramétrico

ajustado aos dados e pode ser vista, para  $z = 6$ , na figura 2.4, representada pela linha preta contínua. Os pontos representam a QLF determinada de uma outra forma pelos autores, e as linhas tracejadas representam ajustes realizados em outro trabalho. Essas outras QLFs não foram incluídas no AddQSO. O ajuste utilizado, sendo uma lei de potência no regime de grandes luminosidades, com uma quebra em baixa luminosidade, é bem representado por uma lei de dupla potência:

$$\Phi(M_{1450}, z) = \frac{10^{k(z-6)} \Phi(M_{1450}^*)}{10^{0,4(\alpha+1)(M_{1450}-M_{1450}^*)} + 10^{0,4(\beta+1)(M_{1450}-M_{1450}^*)}}, \quad (2.4)$$

onde o termo de evolução  $10^{k(z-6)}$  foi incluído para levar em consideração a queda em densidade espacial dos quasares a partir do pico em  $z \approx 2,5$ , um resultado apresentado por Richards et al. (2006). Como a amostra utilizada inclui um intervalo pequeno de *redshift*, o parâmetro de evolução  $k$  foi fixado em  $-0,47$ , conforme resultado derivado por Fan et al. (2001) para a evolução da QLF no regime brilhante, entre  $3 < z < 6$ . Segundo os autores, essa taxa de evolução corresponde a uma diminuição de um fator 2 em densidade espacial na faixa de *redshift* abrangida pela QLF. Ainda, como há poucos objetos mais fracos que a magnitude onde ocorre a quebra em baixa luminosidade, a inclinação da curva do regime fraco foi fixada em  $\alpha = -1,5$ . Na tabela 2.2, listamos todos os valores fixados e os valores melhor ajustado dos parâmetros na equação (2.4).

Tabela 2.2: Constantes e valores de melhor ajuste para a QLF de Willott et al (2010).

$\Phi(M_{1450})^\dagger$	$1,14 \times 10^{-8}$
$M_{1450}^*$	$-25,13$
$k$	$-0,47$
$\alpha$	$-1,5$
$\beta$	$-2,81$

$^\dagger$ QSOs  $\text{Mpc}^{-3} \text{mag}^{-1}$

Conforme foi mencionado, o AddQSO permite que o usuário escolha quais das QLFs ele irá integrar e ainda oferece a opção de fazer uma junção de ambas as funções. Na figura 2.5, mostramos como a opção é selecionada no Portal. Assim como a estrutura do *workflow*, a seleção das configurações também é realizada por uma interface construída através de um código em xml. Para o módulo de integração da QLF, esse código pode ser visto no apêndice A.2. Caso a opção de se utilizar

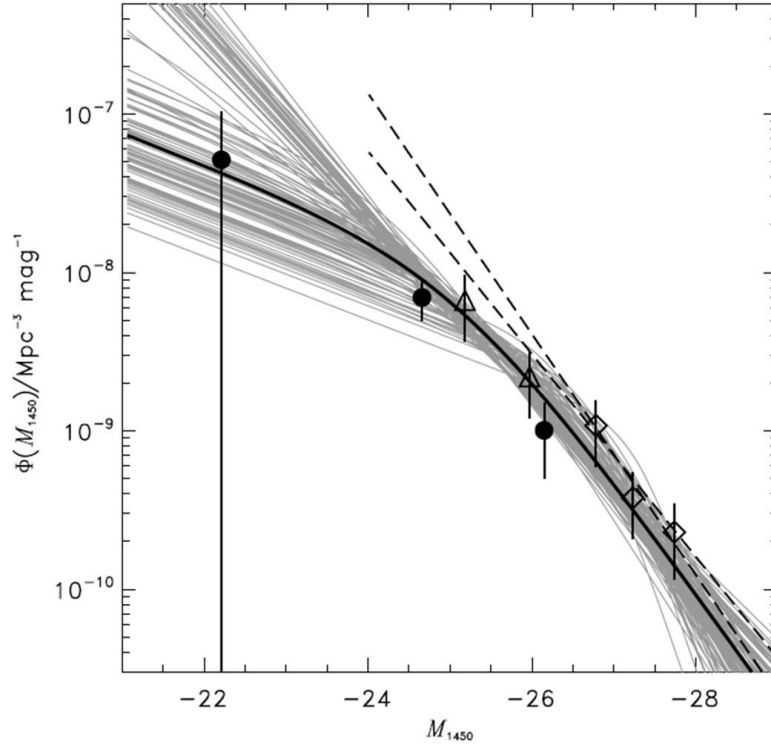


Figura 2.4: QLF de Willott et al (2010) para  $z = 6$ . A função sendo usada está representada pela linha preta contínua e é determinada em função de  $M_{1450}$ .

ambas as QLFs seja selecionada, fica a critério do usuário escolher o *redshift* no qual a mudança entre QLFs será realizada. Como dado de entrada, além da QLF, deve-se prover os parâmetros cosmológicos,  $\Omega_m$ ,  $\Omega_\lambda$ ,  $w$  e  $h$  (que serão utilizados para converter a magnitude absoluta em magnitude aparente), a área total do céu a ser simulada, e os intervalos de  $z$  e magnitude absoluta nos quais a integração será feita. Como o intervalo total de  $z$  representado pela junção das duas QLFs é  $0,3 < z < 6,42$ , qualquer integração feita além desses limites é uma extrapolação das funções e deve ser realizada com cuidado. Isso também é verdade para a junção das duas QLFs, que requer extrapolação de uma ou ambas as funções. Nas figuras 2.6 e 2.7, podem ser observados os parâmetros a serem fornecidos como entrada e seus valores fiduciais.

O AddQSO faz a integração da QLF em intervalos de *redshift*. Para evitar a realização de uma intergração dupla (em *redshift* e em magnitude absoluta), para cada  $\Delta z$  consideramos um valor fixo de  $M_i$  determinado como centro do intervalo de magnitude  $\Delta M$ . Assim, integramos apenas em  $z$ , tendo  $M_i$  como um valor constante. A cada iteração, obtemos o número de quasares em  $\Delta z$  e  $\Delta M$  na área do

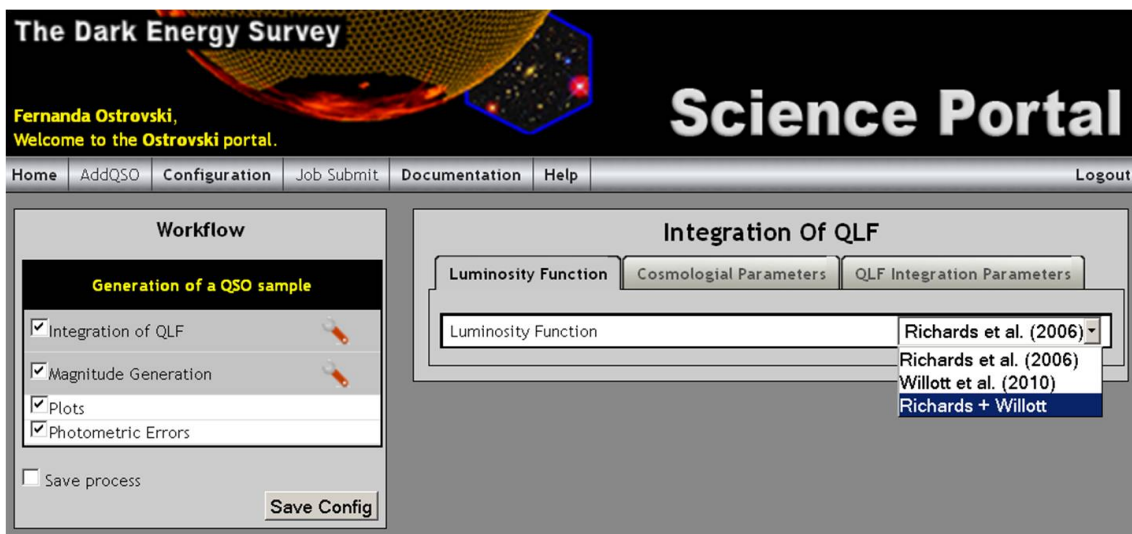


Figura 2.5: Interface do Portal Científico mostrando a aba de seleção da QLF a ser integrada pelo AddQSO.

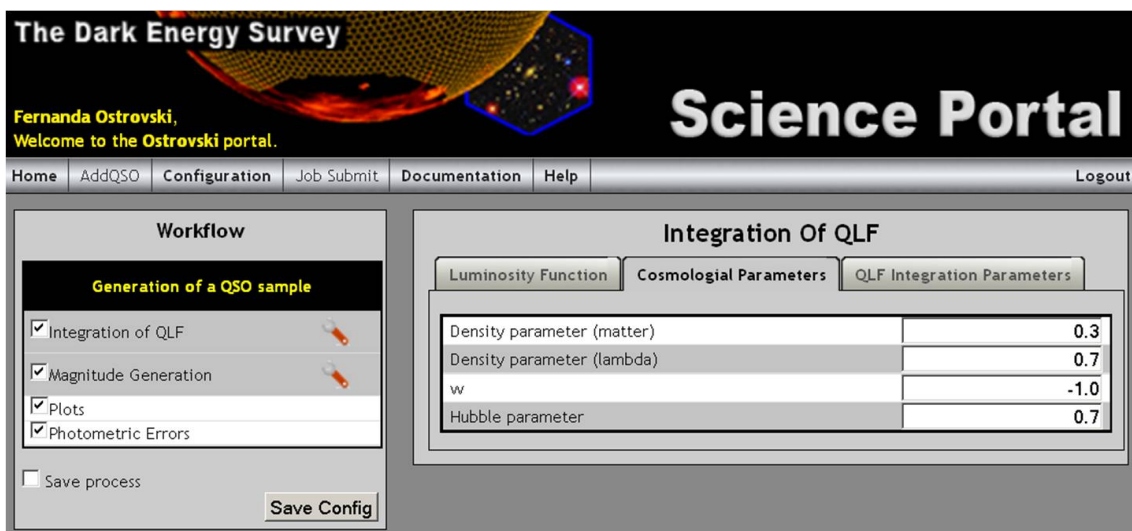


Figura 2.6: Parâmetros cosmológicos fiduciais na aba de configurações da interface AddQSO no Portal Científico.

Integration Of QLF	
QLF Integration Parameters	
Minimum redshift	0.3
QLF change redshift (if Richards + Willott is selected)	5.3
Maximum redshift	8.0
Redshift bin	0.05
Minimum absolute i magnitude (bright)	-30.0
Maximum absolute i magnitude (faint)	-23.5
Absolute magnitude bin	0.05
Simulated area (square degree)	5000.0

Figura 2.7: Parâmetros fiduciais relativos à intergração da QLF pelo AddQSO na aba de configurações da interface do Portal Científico.

céu determinada. Podemos então sortear valores aleatórios dentro desses intervalos de *redshift* e magnitude absoluta, criando quasares simulados. A partir do valor sorteado para  $M_i$  obtemos a magnitude aparente  $m_i$ . No caso da função de Richards et al. (2006), a conversão necessária para obtermos a magnitude absoluta na banda  $i$  em  $z = 0$  é:

$$\begin{aligned} M_i(z = 0) &= M_i(z = 2) + 2,5(1 + \alpha_\nu) \log(1 + 2) \\ &= M_i(z = 2) + 0,596. \end{aligned} \quad (2.5)$$

Já no caso da função do Willott et al. (2010), que se baseia em valores de  $M_{1450}$ , a conversão para  $M_i$  é:

$$\begin{aligned} M_{1450} &= M_i(z = 2) + 0,596 + 2,5\alpha_\nu \log \frac{1450\text{\AA}}{7471\text{\AA}} \\ &= M_i(z = 2) + 1,486. \end{aligned} \quad (2.6)$$

Utilizando a conversão dada pela equação (2.5), temos:

$$M_i(z = 0) = M_{1450} - 0,89. \quad (2.7)$$

Dependendo da área e do intervalo de  $z$  sendo integrado, o resultado pode ser

um número de objetos inferior a um. Nesses casos, o AddQSO aumenta o tamanho do intervalo de integração. A primeira tentativa é aumentar progressivamente o tamanho do intervalo de magnitude absoluta. Esse recurso costuma ser necessário para os intervalos em  $z$  maiores, onde a densidade espacial de objetos é pequena. Caso os limites de  $M_i$  sejam superados e o número de quasares ainda não tenha atingido valores maiores ou iguais a um, o tamanho do intervalo de *redshift* pode ser aumentado. Esse último recurso é exigido apenas quando trabalhamos com grandes *redshifts* em áreas pequenas.

O resultado dessa etapa é um catálogo contendo  $z$ ,  $M_i$  e  $m_i$ . Na tabela 2.3 há um exemplo de um trecho do catálogo que é gerado nessa etapa do AddQSO. Além dos valores de  $z$  e das magnitudes absoluta e aparente, a última coluna apresenta qual a QLF que deu origem àquele objeto. Esse catálogo servirá como entrada no próximo módulo da nossa suíte descrito na sessão a seguir.

Tabela 2.3: Trecho do catálogo de saída do módulo que integra as QLFs.

<i>redshift</i>	$M_i$	$i$	QLF <sup>†</sup>
5,16547657	-22,3138832	26,1133901	1
5,15262621	-22,3193646	26,1015705	1
5,17185107	-22,3120746	26,1183366	1
5,16489686	-22,3205889	26,1063988	1
5,17794601	-22,3351108	26,0982970	1
5,21951535	-31,2962750	17,1574720	2
5,20058390	-29,0535278	19,3909775	2
5,24324747	-28,8239749	19,6413078	2
5,24084768	-28,6621529	19,8019658	2
5,2480135	-28,4030292	20,0645636	2

<sup>†</sup> 1 = Richard et al. (2006); 2 = Willott et al. (2010)

## 2.2 Geração de magnitudes aparentes em outros filtros

O segundo passo na simulação do catálogo é a geração das magnitudes aparentes em outros filtros. No Portal Científico, o usuário pode escolher entre dois conjuntos de filtros:  $g$ ,  $r$ ,  $i$ ,  $z$  e  $Y$  do DES ou  $u$ ,  $g$ ,  $r$ ,  $i$ ,  $z$  do SDSS. Na figura 2.8 são exibidas as curvas de transmissão para os dois conjuntos.

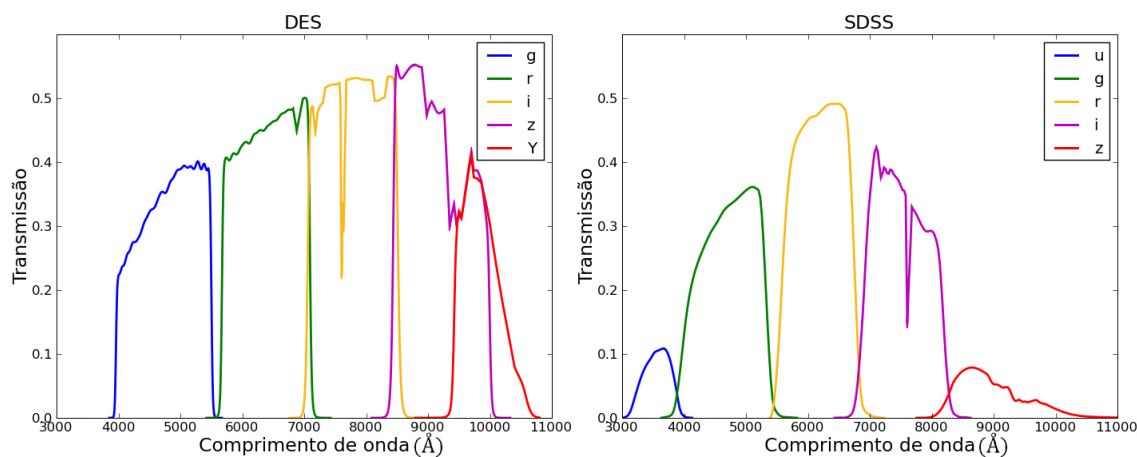


Figura 2.8: Curvas de transmissão dos filtros DES (esquerda) e SDSS (direita) usadas pelo AddQSO.

Além dos filtros, o usuário também pode escolher entre 3 bibliotecas de SEDs, com a possibilidade de combiná-las. Na figura 2.9 mostramos a aba de seleção de configurações dessa etapa do AddQSO no Portal.

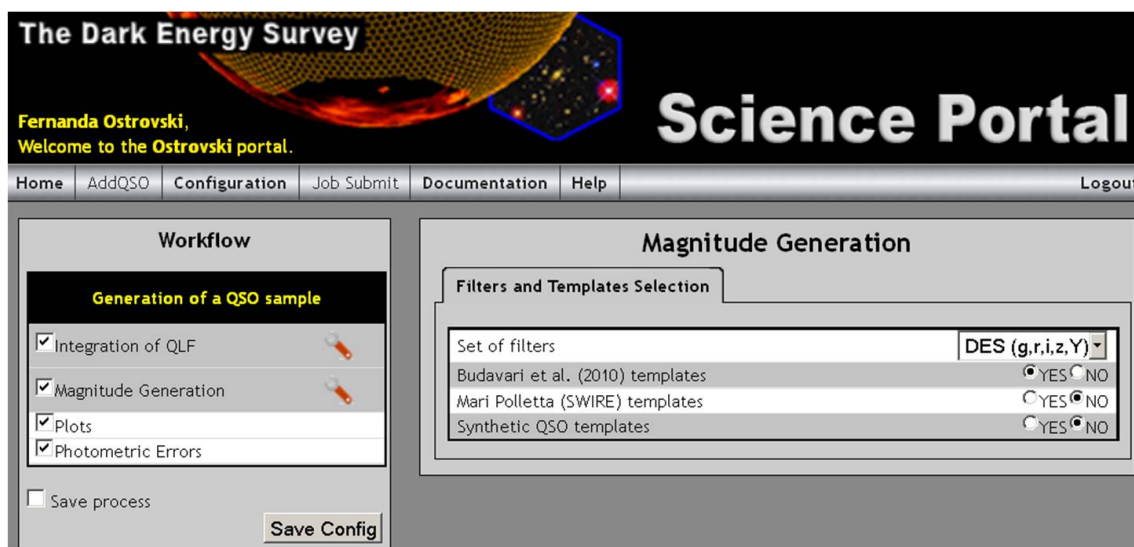


Figura 2.9: Interface do Portal Científico mostrando a aba de seleção do conjunto de filtros e das bibliotecas de SEDs a serem utilizadas pelo AddQSO na geração das magnitudes aparentes em outros filtros.

A primeira biblioteca na lista é a de Budavári et al. (2001). Para criá-la, os autores tentaram separar os quasares presentes em uma amostra de treinamento contendo fotometria e *redshifts* espectroscópicos em diferentes classes. Para isso, eles utilizaram um método próprio para desenvolver um conjunto de modelos espectrais. O resultado é uma biblioteca com oito SEDs que representam dois conjuntos de modelos que diferem na intensidade das linhas de emissão. Na figura 2.10 mostramos



as oito SEDs dessa biblioteca.

As duas bibliotecas restantes foram compiladas do algoritmo de cálculo de *photo-z* LePhare (Arnouts et al. 1999 e Ilbert et al. 2006), que está descrito na sessão 3.2. Uma das bibliotecas, a que nos referimos como "modelos do SWIRE" em referência ao levantamento *Spitzer Wide-Area Infrared Extragalactic Survey* (Lonsdale et al. 2003), são provenientes do trabalho de Mari Polletta (Polletta et al. 2007). No AddQSO, essa biblioteca contém cinco SEDs, sendo três delas representativas de quasares selecionados no óptico com diferentes razões de fluxo entre infravermelho/óptico, e dois quasares tipo 2 (análogo ao que acontece com a subdivisão das galáxias Seyfert, quasares de tipo 2 são AGNs obscurecidos mas de alta luminosidade. A diferença entre eles e os quasares de tipo 1 está na ausência de linhas de emissão largas). Na figura 2.11 podem ser observados todas as SEDs que compõe essa biblioteca. Para o conjunto de tipo 1 (na figura 2.11: QSO1, TQSO1 e BQSO1), os modelos foram obtidos através da combinação entre composição de espectros do SDSS e informações de comprimento de onda de repouso no infravermelho para 35 quasares identificados no SDSS e no SWIRE (Hatziminaoglou et al. 2005). O modelo "QSO2" de tipo 2 foi obtido através da combinação do espectro em óptico e infravermelho próximo observado do quasar avermelhado FIRST J013435.7-093102 (Gregg et al. 2002) e as informações de comprimento de onda de repouso no infravermelho de quasares na amostra do *Palomar-Green* (Polletta et al. 2006). Por último, a SED "Torus" de tipo 2 corresponde ao modelo usado para ajustar a SED do quasar obscurecido SWIRE\_J104409.95+585224.8 (Polletta et al. 2006).

A última biblioteca é composta por espectros sintéticos de quasares. Segundo a descrição do algoritmo LePhare, as SEDs variam em inclinação do contínuo, acentuação das linhas de emissão e componente de corpo negro. No total, há 31 modelos. Na figura 2.12 mostramos uma subamostra dessa biblioteca.

As SEDs em repouso não consideram efeitos de absorção sofridos pela luz do quasar até o observador. A presença de nuvens de hidrogênio neutro do meio intergaláctico faz com que fótons mais energéticos do que o limite de Lyman ( $1216\text{\AA}$ ) sejam absorvidos pelo H I presente na linha de visada do quasar sendo observado (Gunn & Peterson, 1965). Esse efeito, conhecido com efeito de Gunn-Peterson,

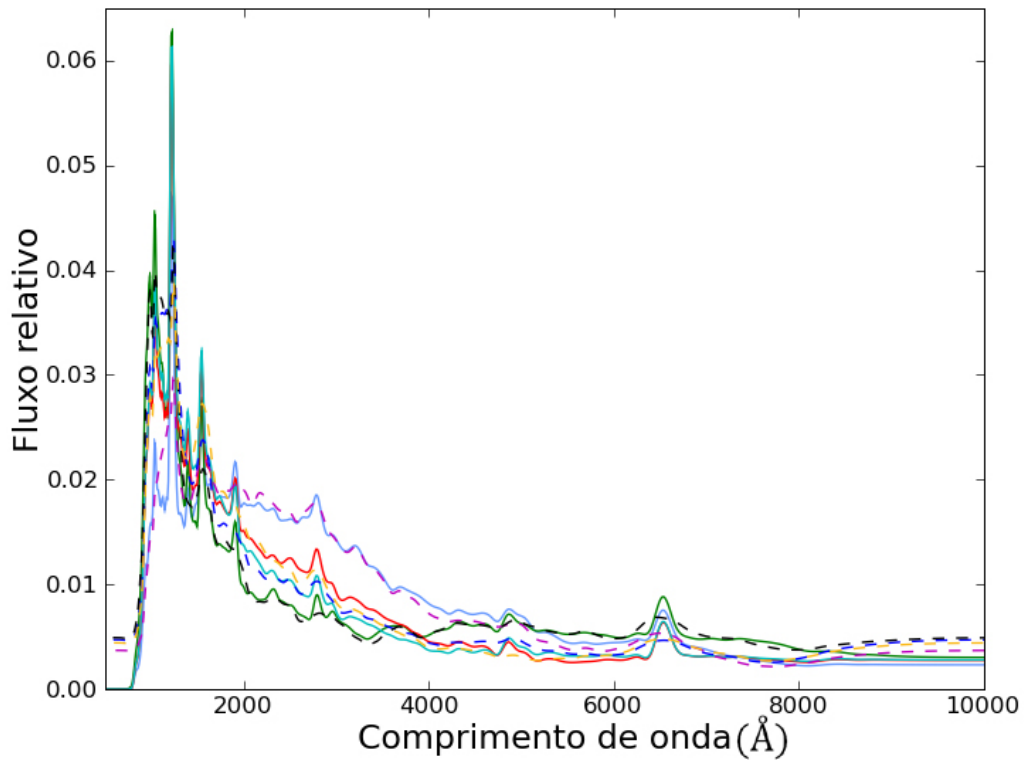


Figura 2.10: SEDs presentes na biblioteca de Budavári et al. (2001). As linhas contínuas e tracejadas distinguem os dois conjuntos de SEDs presentes na biblioteca.

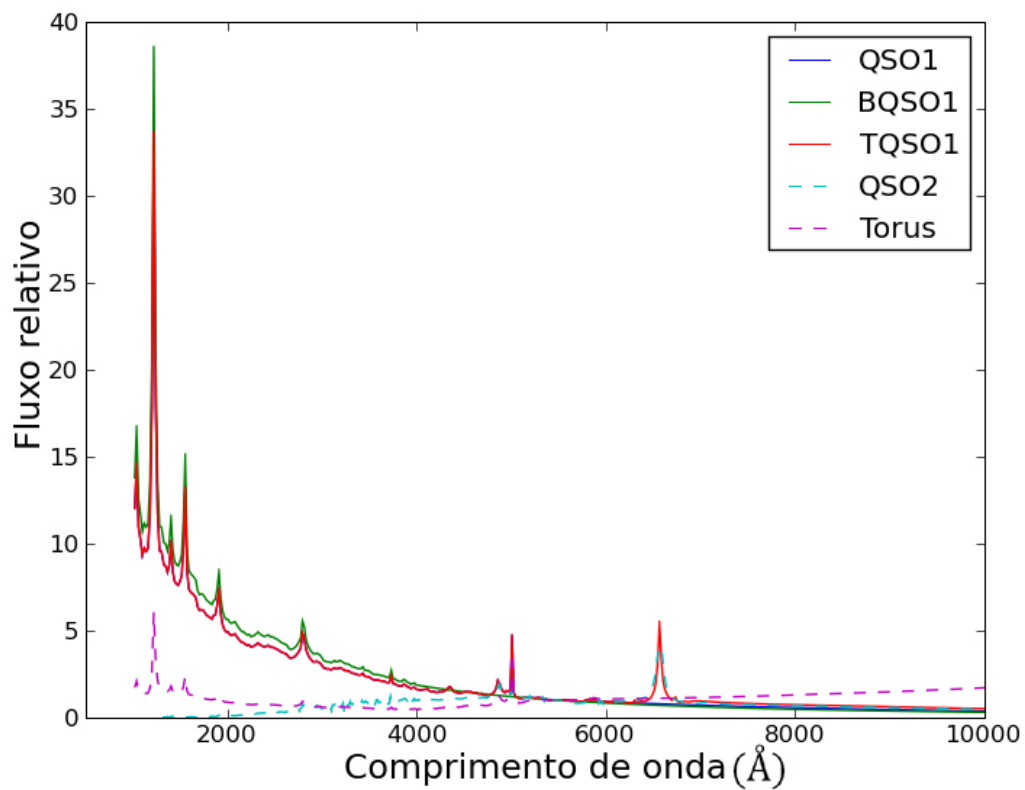


Figura 2.11: SEDs presentes na biblioteca de Mari Polletta (SWIRE). Os modelos representados pelas linhas tracejadas são de quasares tipo 2.

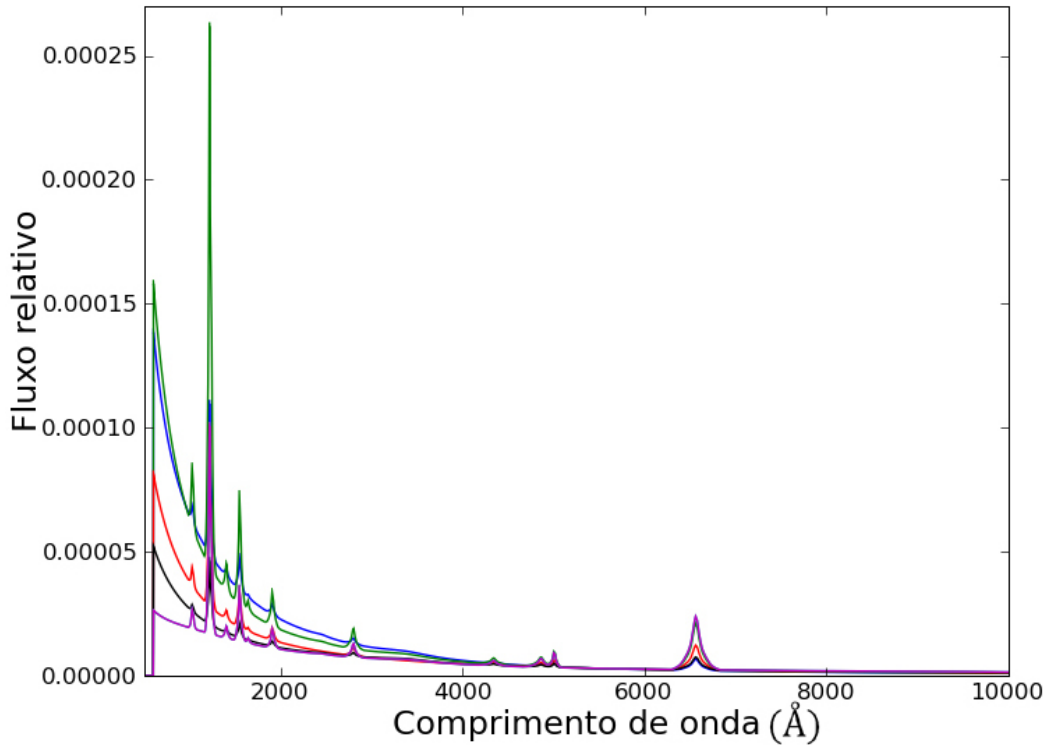


Figura 2.12: Subamostra das SEDs presentes na biblioteca de espectros sintéticos de quasares.

ocasiona uma diminuição do fluxo emitido pelo quasar e precisa ser levado em conta e corrigido nas simulações. A correção se faz necessária para comprimentos de onda menores que  $1216\text{\AA}$  (em repouso), onde se observa a chamada floresta de  $\text{Ly}\alpha$ . Para os filtros do DES em questão, começamos a observar decréscimos na transmissão para fontes que apresentam  $z \geq 2,4$ , quando o efeito se faz presente na banda  $g$ .

Em 2004, Songaila mediu a transmissão da floresta de  $\text{Ly}\alpha$  para observações de 50 quasares com  $2,0 < z < 6,3$ . Baseado em seus resultados, ajustamos o polinômio de 3º grau que representa o fator de transmissão  $f(z)$  em função do *redshift* descrito pela equação:

$$f(z) = 0,02z^3 - 0,30z^2 + 1,0z - 0,20. \quad (2.8)$$

Os dados e o respectivo ajuste podem ser vistos na figura 2.13. Fica claro que, quanto maior o *redshift* do objeto, maior será a absorção da radiação. Para valores de *redshift* maiores que  $z = 6,0$  consideramos o valor de  $f(z) = 0$ .

A redução do fluxo transmitido tem efeitos sobre os valores de magnitude simulados pelo AddQSO em cada filtro para um dado objeto. No caso dos filtros do DES, quasares com  $z > 3,72$  têm suas bandas  $g$  completamente dominadas pela a

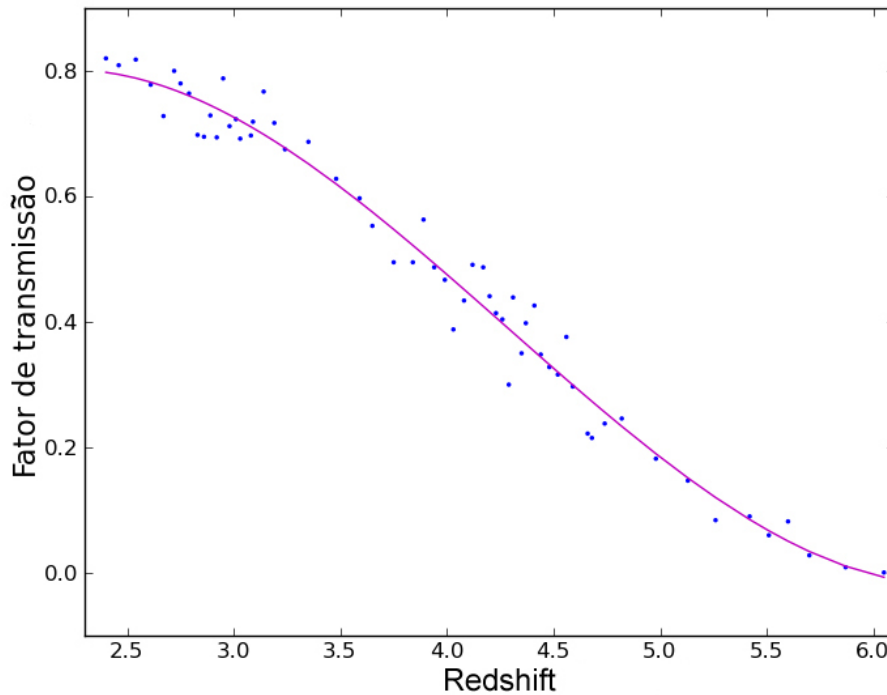


Figura 2.13: Fator de transmissão de acordo com o *redshift*. Os pontos são dados em Songaila (2004) e o ajuste da curva é dado pela equação (2.8).

absorção da floresta de  $\text{Ly}\alpha$ , o que reduz o fluxo observado a pelo menos 50% de seu valor original. Analogamente, objetos simulados com  $z > 5$ , 19 são dominados pelas absorções do  $\text{Ly}\alpha$  na banda  $r$ , reduzindo o fluxo a menos 5% do valores originais. Para o caso da banda  $i$ , a floresta de  $\text{Ly}\alpha$  domina completamente o intervalo de  $\lambda$  quando  $z > 6,72$ . Entretanto, quando  $z = 6,05$ , o fluxo observado de fótons mais energéticos que  $8460\text{\AA}$  (comprimento de onda da  $\text{Ly}\alpha$  desviado para o vermelho) cai a zero. Essa faixa compreende mais da metade da banda  $i$ , o que faz com que quase todos os objetos sejam *i-dropouts* nesse *redshift*.

Conforme foi mencionado, o AddQSO utiliza o catálogo produzido pela integração da QLF como entrada para o módulo que gera as magnitudes aparentes em outras bandas. A escolha da SED é feita de forma diferente dependendo da biblioteca escolhida. Caso a opção selecionada pelo usuário seja usar os 44 modelos presentes nas três bibliotecas, a escolha da SED é não-aleatória e dependente do *redshift*. A distribuição de SEDs nesse caso foi obtida através do ajuste destas aos objetos do catálogo de quasares do DR7 do SDSS (Schneider et al. 2010), a que iremos nos referir como QDR7, pelo LePhare. Consideramos intervalos de 0,5 em *redshift* e a distribuição de SEDs melhor ajustadas ao QDR7 como representativa

do Universo. Mais informações sobre esse ajuste serão fornecidas na seção 3.2. Se o usuário escolher outra combinação de bibliotecas, a seleção de SEDs é aleatória. Independente de como ocorre a seleção do modelo, o *redshift* do objeto é aplicado aos comprimentos de onda de repouso da SED e uma *spline* é ajustada aos novos pontos. As magnitudes são determinadas através de suas relações com os fluxos nos respectivos filtros:

$$m_x = m_0 - (2,5 \log F_x) + C, \quad (2.9)$$

onde  $x$  é o filtro em questão,  $m_0$  é a magnitude de ponto zero, que irá garantir magnitudes coerentes independente das unidades de fluxo das SEDs sorteadas para cada objeto e  $C$  é um fator de correção para obter as magnitudes no sistema AB. Esse fator foi determinado como o valor necessário para que a estrela Vega apresentasse as magnitudes no sistema AB em cada um dos filtros. Para tal, utilizamos o AddQSO e o template da estrela para gerar as magnitudes a partir de  $m_i$ . Já para determinar  $m_0$ , utilizamos a magnitude aparente na banda  $i$ , tal que:

$$m_0 = m_i + (2,5 \log F_i). \quad (2.10)$$

Isso é feito através da convolução da SED com o filtro  $i$  e posterior cálculo da área sob a curva, o que equivale ao fluxo integrado em  $i$ . Uma vez determinada  $m_0$ , podemos aplicar à SED, se preciso, a correção da floresta de  $\text{Ly}\alpha$ , convoluir a SED corrigida com os outros filtros, calcular os fluxos integrados e determinar as magnitudes.

Nessa etapa, não aplicamos nenhum corte às magnitudes geradas. Atribuímos valores 99,999 apenas a casos em que o fluxo integrado for zero, o que ocorre devido à atenuação imposta pelo acréscimo dos efeitos da floresta de  $\text{Ly}\alpha$ . Programacionalmente, esse valor é tratado como NaN, o que significa "not a number". Normalmente isso ocorre quando há uma tentativa de divisão por zero, ou, no caso do AddQSO, quando se tenta calcular o logaritmo de zero.

O catálogo gerado no fim dessa etapa contém *redshift* e magnitudes aparentes nas bandas selecionadas. Na tabela 2.4 exibimos como exemplo um trecho de um catálogo gerado com magnitudes SDSS. A primeira coluna traz uma identificação para cada objeto. A última coluna identifica a SED utilizada para gerar as magnitudes.

Tabela 2.4: Trecho do catálogo de saída do módulo que gera as magnitudes aparentes. No caso, as magnitudes foram geradas nos filtros SDSS.

id	<i>redshift</i>	<i>u</i>	<i>g</i>	<i>r</i>	<i>i</i>	<i>z</i>	SED
11778209	3,87540170	25,363	24,400	24,264	24,290	24,330	16
11778210	3,87732620	25,509	24,246	24,283	24,402	24,486	14
11778211	3,85199805	25,470	24,213	24,254	24,373	24,478	14
11778212	3,87571934	25,528	24,366	24,246	24,267	24,304	26
11778213	3,89167851	25,176	24,345	24,259	24,305	24,342	17
11778214	3,85451211	25,335	24,375	24,248	24,274	24,319	16
11778215	3,86581127	25,331	24,369	24,237	24,263	24,306	16
11778216	3,89186912	25,399	24,434	24,290	24,316	24,351	16
11778217	3,86623824	25,492	24,232	24,270	24,389	24,483	14
11778218	3,88123439	25,213	24,268	24,257	24,336	24,390	20
11778219	3,87441192	21,605	24,299	24,262	24,284	24,339	09

### 2.3 Atribuição de erros para as magnitudes

Apesar da simulação estar completa após a geração das magnitudes aparentes, só é possível utilizar o catálogo para testar ferramentas de análise, classificação de objetos e cálculo de *redshifts* fotométricos na presença de erros fotométricos. Para gerá-los, nos baseamos na distribuição de erros das estrelas simuladas pela colaboração do DES em uma amostra do chamado *Data Challenge 6B* (DC6B). O DC6B simulou exposições em um conjunto de 62 CCDs representando o equipamento da DECam em dez noites nas cinco bandas do DES. As exposições foram coadicionadas originando uma tabela com  $\approx 2,4 \times 10^7$  objetos entre estrelas e galáxias (quasares não foram incluídos nessa versão da simulação) em uma área de mais de 140 graus quadrados.

Para atribuímos erros fotométricos aos QSOs, nos baseamos numa sub-amostra de  $5 \times 10^5$  estrelas. Para evitar contaminação por galáxias, que, sendo objetos extensos não apresentam a mesma distribuição de erros, selecionamos apenas as estrelas classificadas como tal na tabela verdade da simulação. Outras exigências foram: as magnitudes verdadeiras (de entrada) devem estar entre 14,0 e 28,0; a diferença absoluta entre a magnitude verdadeira e a simulada deve ser inferior a 3,0 mag; o erro na magnitude simulada deve ser inferior a 2,0 mag. Todos estes limites foram aplicados para cada banda de forma que, por exemplo, limites impostos sobre a banda *i* não influenciassem a geração de erros na banda *g*. As magnitudes e os erros são de tipo "mag\_auto", ou seja, são magnitudes de abertura automática, calculadas

pelo algoritmo SExtractor<sup>1</sup> (Bertin & Arnouts 1996), que constrói catálogos a partir de imagens astronômicas. No Portal, a geração de um catálogo contendo erros, por enquanto, só é possível para as simulações com filtros DES.

Antes de distribuímos os erros fotométricos, precisamos considerar efeitos da observação nas magnitudes simuladas pelo AddQSO. Idealmente, os quasares seriam observados conforme as contagens produzidas pela simulação. No entanto, o AddQSO não leva em consideração efeitos gerados pela instrumentação no momento da observação, ou consequências da redução de dados. Sendo assim, acrescentamos dois efeitos aos quasares simulados: uma diferença entre as magnitudes simulada e observada e os erros fotométricos em si. No painel da esquerda da figura 2.14 podem ser observadas as diferenças entre a magnitude simulada e a magnitude observada ( $\Delta m$ ) em função da magnitude simulada das estrelas da amostra do DC6B para a banda  $i$ . Sobre os pontos estão representadas as médias e desvios padrão para cada um dos intervalos de 0,1 mag, calculados após a remoção dos pontos que estavam fora do limite de  $3\sigma$ . Para atribuir esses valores de  $\Delta m$  para os quasares, fazemos um sorteio para uma distribuição gaussiana com os valores obtidos para cada intervalo. No painel da direita da figura 2.14 temos a distribuição de  $\Delta m$  gerada pelo AddQSO para uma simulação de quasares com magnitudes nos filtros DES e 5000 graus quadrados de área. As distribuições de  $\Delta m$  das estrelas e dos quasares simulados para as outras bandas podem ser vistas no apêndice B.1.

A geração dos erros para as magnitudes dos quasares é feita com base na relação entre o erro e a magnitude observada, que pode ser vista para as estrelas do DC6B na banda  $i$  no painel da esquerda da figura 2.15. Sobre os pontos estão representados a mediana e o desvio padrão dos dados. A atribuição dos erros a cada objeto simulado pelo AddQSO é feita seguindo uma gaussiana centrada na mediana e com o respectivo desvio padrão para cada intervalo de 0,1 mag. Uma vez que a distribuição de erros do DC6B não segue exatamente uma forma gaussiana, limitamos o valor máximo possível de ser atribuído como erro acima da mediana em  $1,5\sigma$  e o valor mínimo abaixo da mediana em  $1,1\sigma$  para que fosse mantido o comportamento geral da distribuição de erros das estrelas do DC6B. O painel da direita da figura 2.15 mostra a distribuição de erros obtidas com o AddQSO. As

---

<sup>1</sup><http://terapix.iap.fr/soft/sExtractor/>

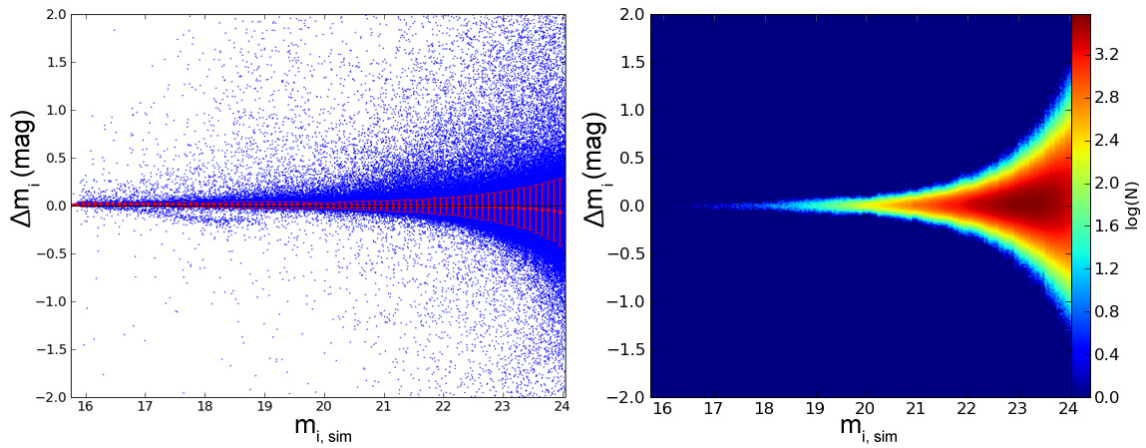


Figura 2.14: Distribuição de  $\Delta m_i$  em função da magnitude simulada na banda  $i$ . No painel da esquerda vemos em azul a distribuição de  $\Delta m_i$  para as estrelas do DC6B e, sobrepostas em vermelho, as médias com barras de  $1\sigma$  calculadas após a remoção dos objetos que apresentavam  $\Delta m_i$  fora do limite de  $3\sigma$ . No painel da direita está a distribuição de  $\Delta m_i$  simulada para quasares em uma área de 5000 graus quadrados no filtro  $i$  do DES. A barra colorida representa o logaritmo da contagem de objetos em cada ponto.

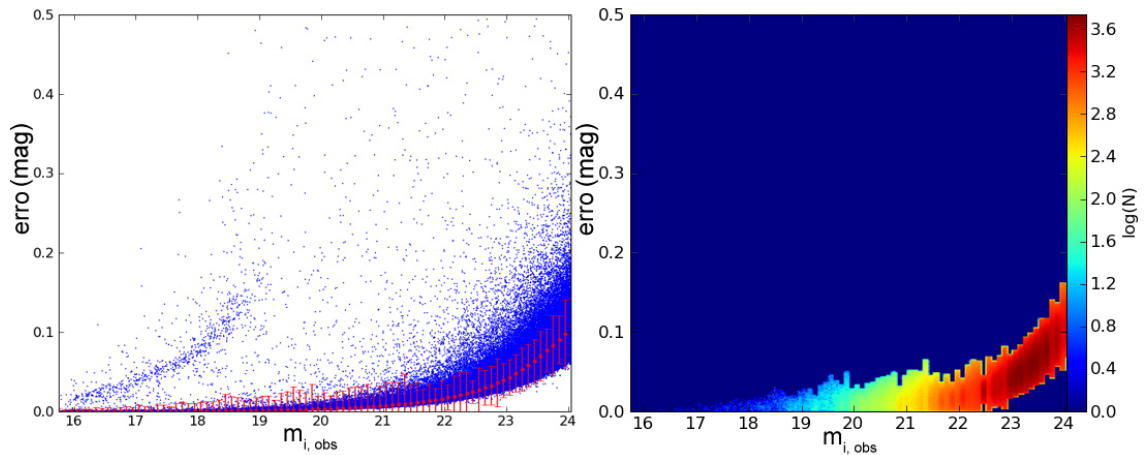


Figura 2.15: Erros fotométricos na banda  $i$  em função da magnitude observada na mesma banda. No painel da esquerda vemos em azul a distribuição de erro para as estrelas do DC6B e, sobrepostas em vermelho, as medianas e as barras de  $1\sigma$ . No painel da direita está a distribuição de erros fotométricos simulada para quasares em uma área de 5000 graus quadrados no filtro  $i$  do DES. A barra colorida representa o logaritmo da contagem de objetos em cada ponto.



distribuições de erros em função da magnitude observada nas outras bandas são apresentadas no apêndice B.2.

## 2.4 Distribuição espacial

A distribuição espacial dos objetos simulados aqui descrita é um processo simplificado. Idealmente, os QSOs deveriam ocupar posições de acordo com a distribuição de halos de matéria escura do Universo. Contudo, não há um consenso na literatura em relação a uma função de correlação que caracterize essa distribuição e que possa ser assimilada pelo nosso algoritmo de simulação. Tentativas de distribuir nossos objetos de acordo com a simulação de halos realizadas pela colaboração do DES não puderam ser concluídas devido à resolução da simulação fazer com que os catálogos de halos sejam mais rasos que os catálogos de QSOs para  $z > 1,0$ . Sendo assim, os QSOs são distribuídos aleatoriamente no céu.

## 2.5 Validação dos algoritmos

Para garantir o funcionamento do AddQSO e sua capacidade de produzir resultados compatíveis com a realidade, geramos um catálogo de quasares usando o conjunto de filtros do SDSS. O catálogo, a que iremos nos referir como SDSS\_8417, foi gerado através da integração da QLF de Richards et al. (2006) entre  $0,3 < z < 5,5$ . Portanto, estendemos um pouco os limites dessa QLF. Um resumo das características gerais do catálogo pode ser visto na tabela 2.5.

Tabela 2.5: Características do catálogo SDSS\_8417.

Nome	Área <sup>†</sup>	N <sup>o</sup> de QSOs <sup>††</sup>	QLF <sup>‡</sup>	$z_{min}$	$z_{max}$	$M_{i,min}$	$M_{i,max}$	SEDs
SDSS_8417	8417	16603122	1	0,3	5,5	-23,0	-31,0	44

<sup>†</sup>em graus quadrados

<sup>††</sup>sem cortes em magnitude

<sup>‡</sup>Richards et al. 2006

Infelizmente, não existe uma amostra de quasares espectroscopicamente confirmados que não sofra com os vieses de seleção. No caso do SDSS utilizamos, para comparar as cores dos quasares simulados, o catálogo obtido com dados do QDR7. A distribuição de *redshifts* desse catálogo pode ser vista na figura 2.16. Assim como

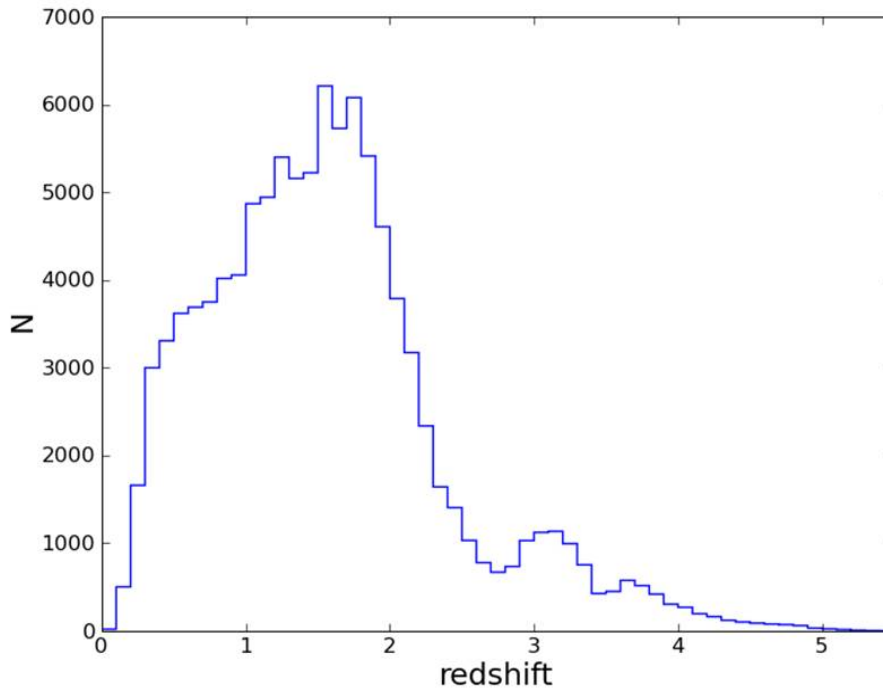


Figura 2.16: Distribuições de *redshift* para os  $\approx 105000$  quasares do catálogo QDR7 (Schneider et al. 2010).

aconteceu com a amostra de Richards et al. (2006), há uma baixa eficiência de seleção de objetos nos *redshifts*  $z = 2,7$  e  $z = 3,5$ , o que ocasiona a baixa contagem de objetos nesses *redshifts*.

Na figura 2.17 comparamos as cores dos quasares simulados pelo AddQSO com as cores dos quasares do QDR7. Limitamos o catálogo SDSS\_8417 aos mesmo limites observados no catálogo do DR7:  $m_u \leq 26,78$ ,  $m_g \leq 26,42$ ,  $m_r \leq 22,88$ ,  $m_i \leq 22,36$  e  $m_z \leq 22,65$  que, é importante ressaltar, estão acima dos limites de magnitude do SDSS<sup>2</sup>. Os caminhos discretos dos objetos simulados são decorrentes da falta de erros fotométricos na nossa amostra idealizada. Apesar disso, há uma correlação entre as cores das amostras, com algumas excessões. Nesses diagramas, representamos por cores diferentes os quasares do catálogo SDSS\_8417 de acordo com a biblioteca de SEDs que gerou as magnitudes aparentes de cada objeto. Para ajudar a identificar cada conjunto de SEDs, normalizamos uma amostragem de cada grupo e as mostramos em um mesmo diagrama na figura 2.18, mantendo as mesmas cores dos quasares do SDSS\_8417 nos diagrams cor-cor.

Os objetos do catálogo SDSS\_8417 com índices de cor  $u - g < -0,5$  e  $g - r \approx$

<sup>2</sup>Os limites de magnitude do SDSS são  $m_u = 23,0$ ,  $m_g = 22,2$ ,  $m_r = 22,2$ ,  $m_i = 21,3$  e  $m_z = 20,5$

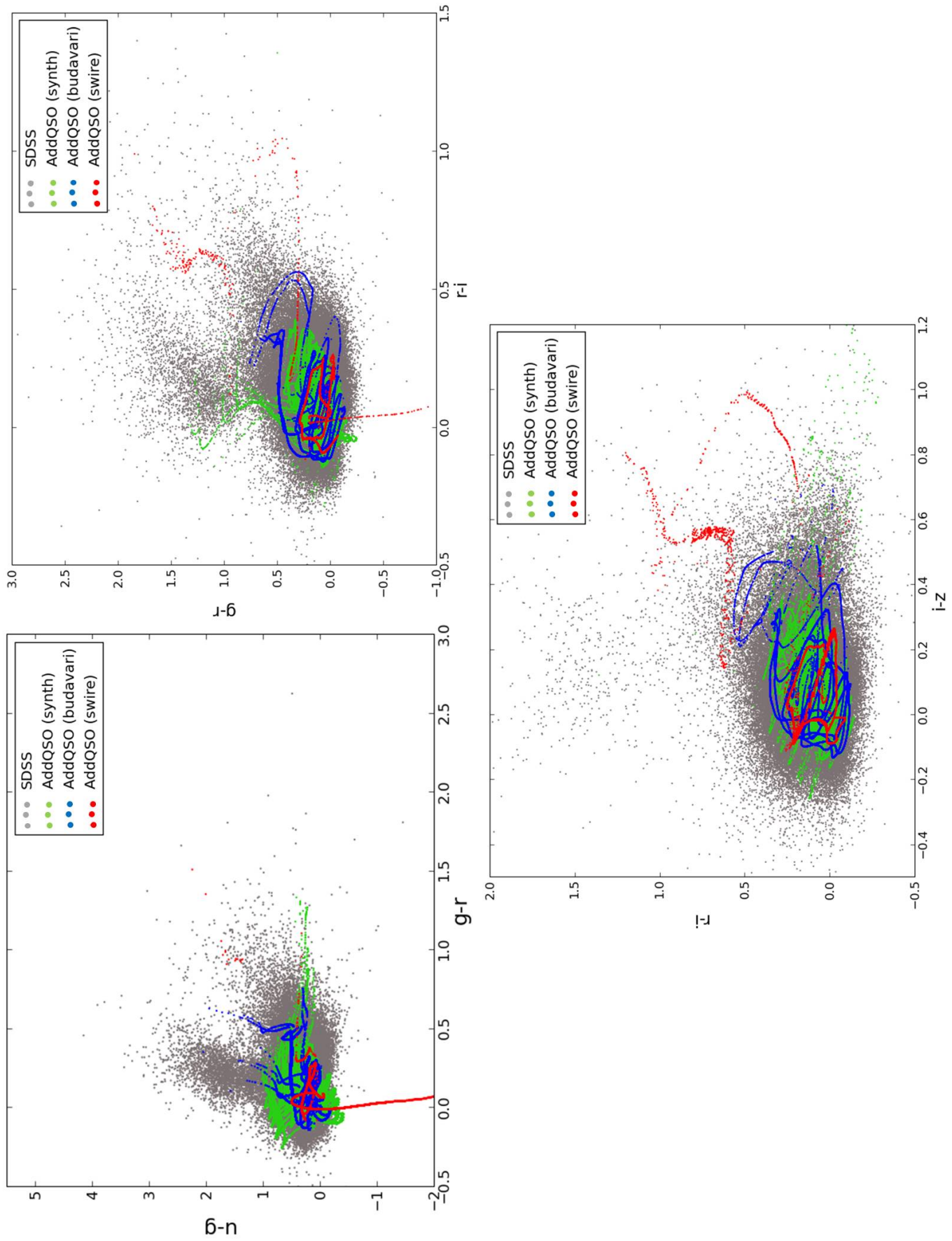


Figura 2.17: Comparação entre os diagramas cor-cor dos quasares do catálogo do QDR7 (cinza) e dos quasares simulados pelo AddQSO. As diferentes bibliotecas de SEDs estão codificadas pelas cores indicadas nos painéis.

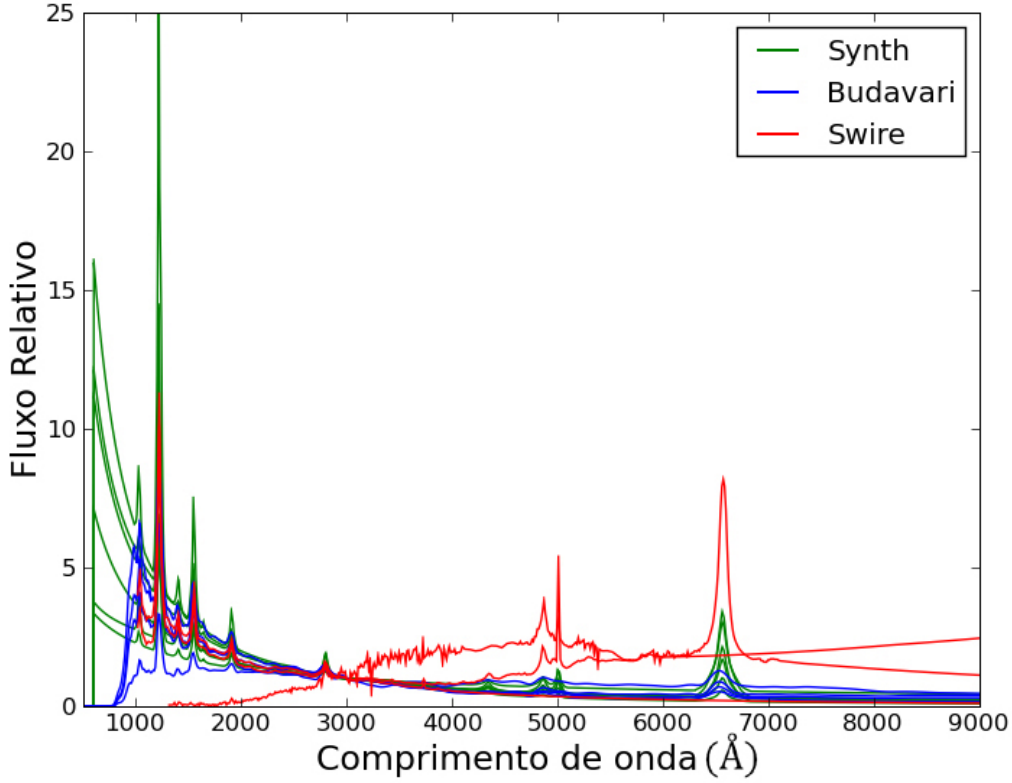


Figura 2.18: Amostragem de SEDs de cada uma das bibliotecas presentes no AddQSO utilizadas na geração das magnitudes aparentes do catálogo SDSS\_8417. As SEDs foram normalizadas em 3000 Å. As cores representam o mesmo que na figura 2.17.

0,0 que aparecem no diagrama que mostra  $u - g \times g - r$  tiveram suas magnitudes geradas a partir de apenas duas SEDs, ambas contidas na biblioteca SWIRE e representativas de quasares de tipo 1. Eles representam apenas 0,4% da amostra. Esses objetos são os mesmos que apresentam índices de cor  $g - r < -0,2$  e  $r - i \approx 0.1$  no diagrama que mostra  $g - r$  em função de  $r - i$ . Mais da metade dos objetos apresenta *redshift* em torno de 2,5, com o restante tendo  $2,6 \leq z \leq 4,0$ . Apesar de se desviarem do comportamento geral da distribuição de cor dos quasares do QDR7, é importante lembrar que a amostra de quasares reais foi submetida a um processo de seleção que reduziu o número de objetos nessa faixa de *redshift* e ainda aplicou limitações em cor. Sendo assim, é possível que a falta desses objetos na distribuição de cores dos quasares do QDR7 seja um efeito de seleção.

Outro ponto de discrepância entre as distribuições de cor dos quasares simulados e dos reais é a falta de objetos que apresentam  $u - g > 1,0$  e  $0,0 < g - r < 0,4$  no diagrama de  $u - g \times g - r$  e de objetos que apresentam  $r - i > 0,6$  e  $-0,2 < i - z < 0,4$  no diagrama de  $r - i \times i - z$ . Esses objetos apresentam *redshifts* superiores a 3,5

e com isso têm, em geral, magnitudes aparentes acima do limite do SDSS. Isso significa que na maior parte dos casos essas magnitudes foram obtidas de outras formas ou tem grandes erros associados e não podem sofrer uma comparação direta com os quasares simulados pelo AddQSO. De fato, se relaxarmos o limite de magnitude imposto ao catálogo SDSS\_8417, objetos começam a preencher essas regiões.

Testamos também a evolução das cores em função do *redshift*. Na figura 2.19 mostramos a média e o desvio padrão de cada cor separadas em intervalos de *redshift*. Em geral, as cores dos quasares do QDR7 (triângulos) e do catálogo SDSS\_8417 (círculos) concordam dentro de suas barras de erro. Contudo, elas são mais discrepantes com o aumento do *redshift*. Isso é de se esperar devido à pequena amostragem de objetos com *redshift* superior a 3,0 no catálogo de quasares QDR7. O único ponto do catálogo SDSS\_8417 que diverge claramente dos quasares reais é o que representa os objetos com *redshift* médio igual a 3,5, que, refletindo a presença dos objetos com índice de cor  $u-g < -0,5$ , se mostra bastante deslocado e com uma grande barra de erro. Nos diagramas de cor  $u-g \times g-r$  e  $r-i \times i-z$ , os quasares do catálogo QDR7 apresentam um intervalo de *redshift* a mais. No caso dos quasares do SDSS\_8417, os objetos eram *dropouts* em pelo menos uma das bandas do diagrama, o que não tornou possível a determinação do locus. A diferença também pode ser atribuída à obtenção das magnitudes aparentes desses objetos na amostra de quasares reais.

Na falta de uma amostra completa espectroscopicamente confirmada, utilizamos a amostra de um milhão de quasares não-confirmados de Richards et al. (2009), a que iremos chamar de QR09, para compararmos as contagens em cada banda. Esse catálogo é composto por  $\approx 1,1 \times 10^6$  candidatos a quasares selecionados fotometricamente de dados do SDSS. A magnitude limite na banda  $i$  é de 21,3 e a área total é de 8417 graus quadrados. Os autores esperam que pelo menos  $8,5 \times 10^5$  dos objetos do catálogo sejam de fato quasares. Como o catálogo não contém *espec-z*, não é possível comparar a distribuição de *redshift*. Entretanto, podemos comparar as contagens na banda  $i$  desse catálogo com as do SDSS\_8417, conforme pode ser visto na figura 2.20. Para essa comparação, limitamos o catálogo SDSS\_8417 em  $m_i \leq 21,3$ , não impondo limites sobre as outras magnitudes, conforme foi feito com a seleção de candidatos do catálogo QR09. No entanto, a seleção de candidatos é feita através de um método que exige treinamento a partir de uma amostra espec-

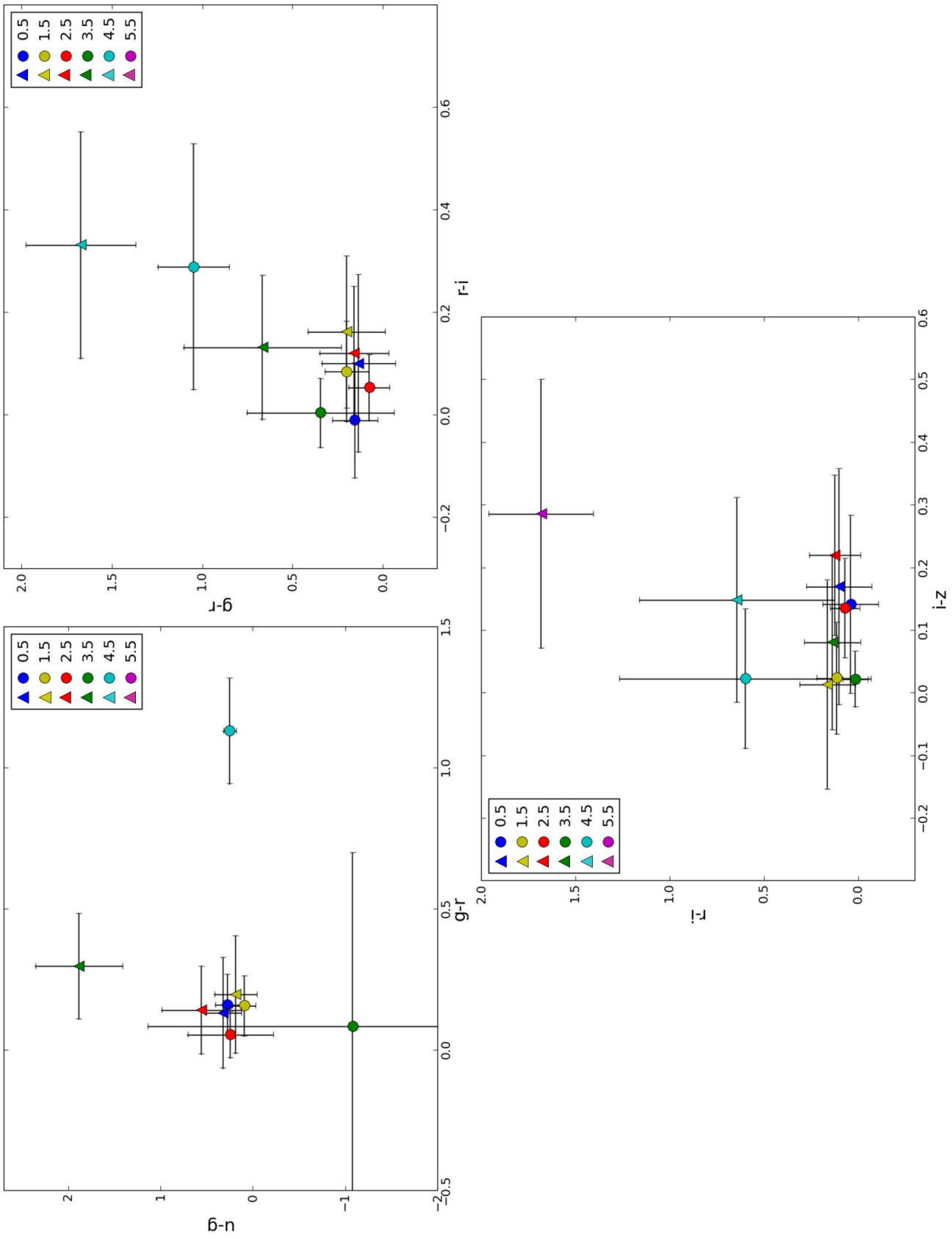


Figura 2.19: Cores médias por intervalo de 1,0 em *redshift*. As cores representam o centro de cada intervalo. Os círculos mostram os valores para os quasares SDSS\_8417 e os triângulos denotam as médias do catálogo de quasares QDR7.

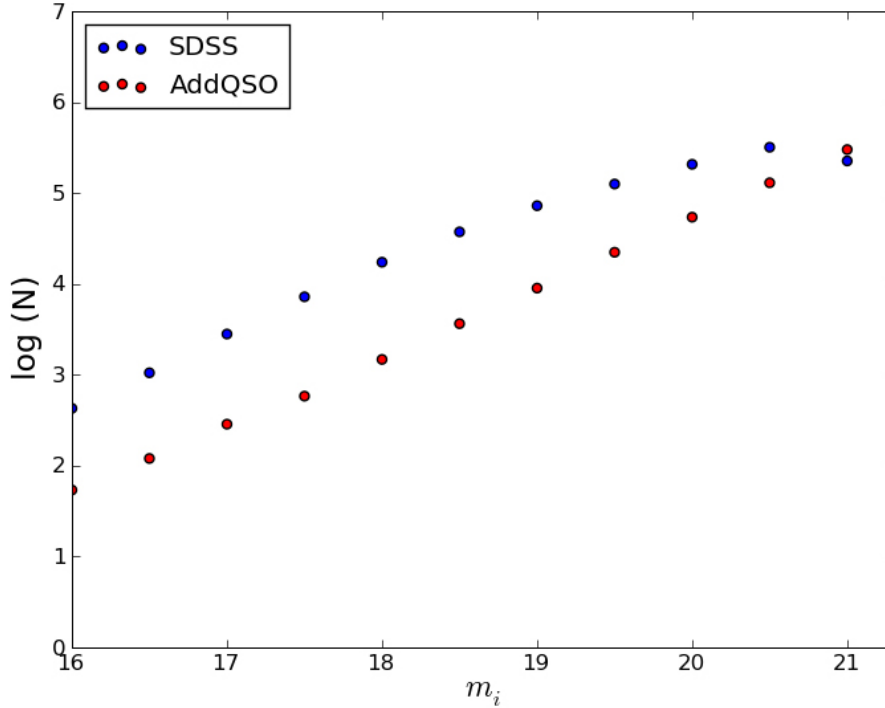


Figura 2.20: Contagens na banda  $i$  para o catálogo QR09 de  $\approx 1 \times 10^6$  candidatos a quasares do SDSS (azul) e o catálogo de quasares simulados SDSS\_8417 (vermelho).

troscópica, o que pode ter levado a vieses nas distribuições de cor e nas contagens.

Pela figura, o AddQSO parece subestimar a quantidade de quasares com magnitudes aparentes fortes na banda  $i$ . No entanto, é preciso levar em consideração que a amostra do QR09 contém contaminação por outros objetos como estrelas e outros tipos de AGN que podem corresponder a uma parte dessa população. Além disso, as magnitudes utilizadas por essa amostra são *über-calibrated* e não medimos o impacto dessa calibração.

## 2.6 Previsões para o DES

Para fazermos previsões sobre o número de quasares esperados na área do DES, sua distribuição de *redshifts* e suas cores, simulamos um catálogo em uma área de 5000 graus quadrados. Iremos nos referir a esse catálogo com DES\_5000 e um resumo de suas principais características pode ser visto na tabela 2.6.

A distribuição de *redshifts* sem cortes em magnitude é exibida na figura 2.21. As linhas vermelhas mostram a dificuldade de juntar QLFs distintas caracterizada pela diferença na inclinação da distribuição. Se estendermos a QLF de Richards

Tabela 2.6: Características o catálogo DES\_5000.

Nome	Área <sup>†</sup>	Nº de QSOs <sup>††</sup>	$z_{min}$	$z_{max}$	$z_{join}$ <sup>‡</sup>	$M_{i,min}$	$M_{i,max}$	SEDs
DES_5000	5000	11928502	0,3	8,0	5,25	-22,9	-31,0	44

<sup>†</sup>em graus quadrados

<sup>††</sup>sem cortes em magnitude

<sup>‡</sup>*redshift* onde as duas QLFs foram unidas

et al. (2006) para *redshifts* maiores, iremos subestimar a amostra de quasares de grandes *redshifts*, e não seremos capazes de simular quasares acima de  $z \approx 6,5$ . Por outro lado, a extensão da QLF de Willott et al. (2010) para *redshifts* menores não iria representar o pico da distribuição de quasares em  $z \approx 2,3$ , subestimando a quantidade de objetos entre  $1,0 < z < 5,0$  e superestimando os quasares próximos com  $z < 1,0$ .

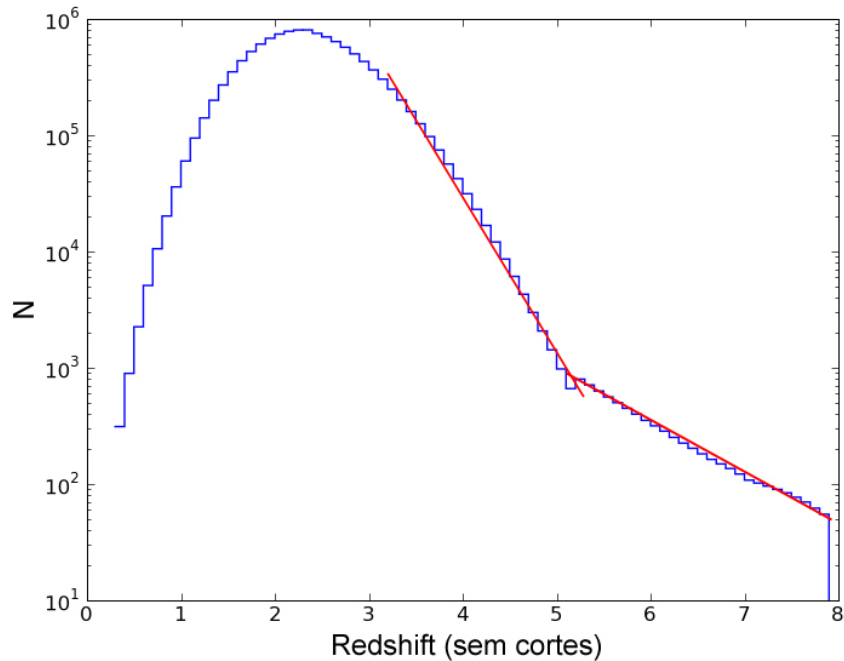


Figura 2.21: Distribuição de *redshifts* dos quasares do catálogo DES\_5000 sem cortes em magnitude. As linhas vermelhas representam as tendências de cada QLF usada na simulação caso elas fossem estendidas muito além de seus regimes de validade em *redshift*.

Se limitarmos o catálogo DES\_5000 em  $m_g < 25,21$ ,  $m_r < 24,77$ ,  $m_i < 24,04$ ,  $m_z < 23,44$ ,  $m_Y < 21,70$ , que são os limites de magnitude esperados pelo DES, e considerarmos apenas os quasares que contém pelo menos uma das bandas dentro dos limites, a distribuição de *redshift* muda e pode ser vista na figura 2.22. Com os erros fotométricos incluídos, as contagens em cada banda são apresentadas na figura



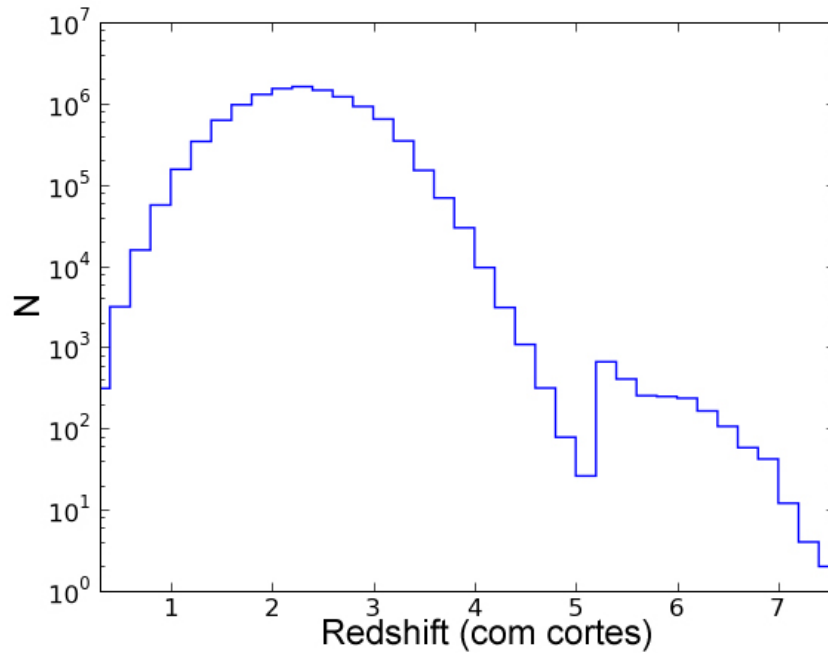


Figura 2.22: Distribuição de *redshift* dos quasares do catálogo DES\_5000 que contém pelo menos uma das magnitudes dentro dos limites  $m_g < 25,21$ ,  $m_r < 24,77$ ,  $m_i < 24,04$ ,  $m_z < 23,44$ ,  $m_Y < 21,70$ .

### 2.23.

Na figura 2.24, mostramos os diagramas cor-cor, onde são exibidos os objetos que têm as três bandas em questão em cada diagrama dentro dos limites impostos. Também exibimos a evolução com o *redshift* dos objetos nos diagramas cor-cor. Os limites de magnitude e a presença de *dropouts* definem o *redshift* máximo em cada cor. Calculamos as cores médias e seus desvios padrão para diferentes intervalos de *redshift*, cujo resultado pode ser visto na figura 2.25.

Se considerarmos um quasar observável pelo DES como aquele que tem pelo menos uma das bandas dentro dos limites de fotometria que impusemos, o total de quasares visíveis é 11668480. O número de objetos por banda por intervalo de *redshift* pode ser visto na tabela 2.7.

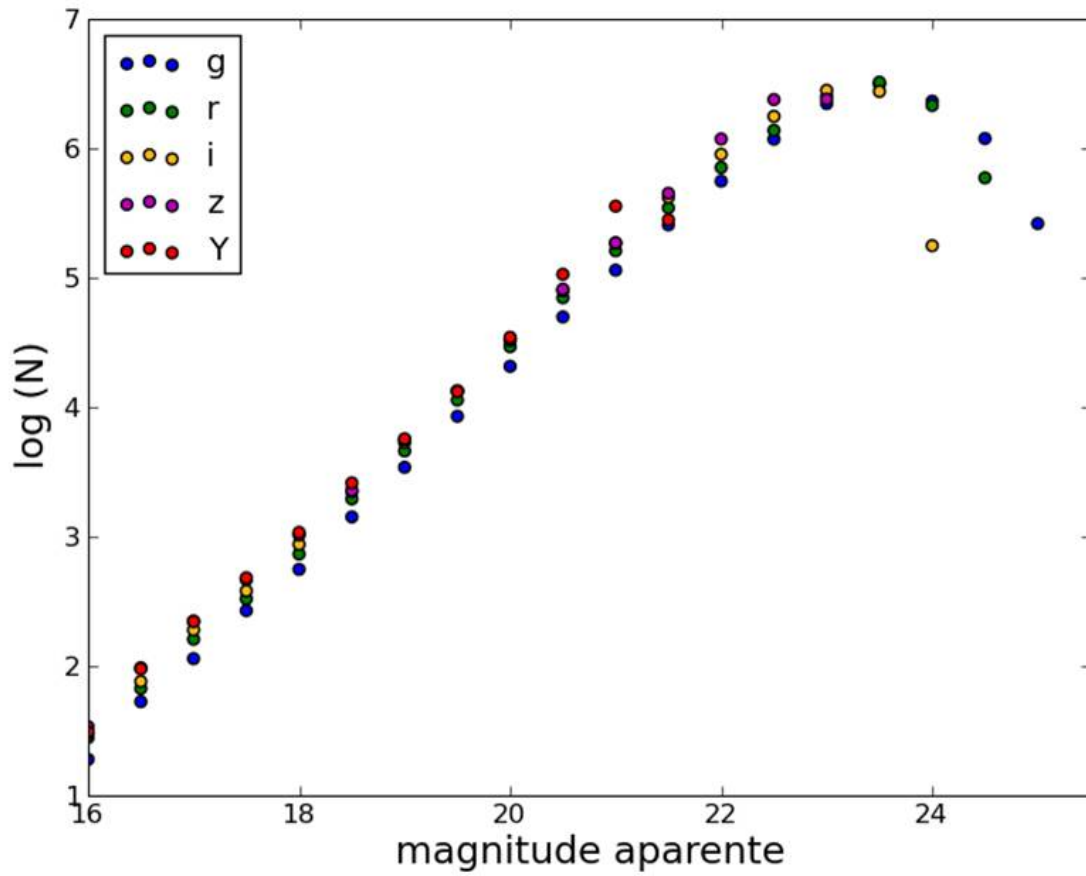


Figura 2.23: Contagens esperadas em cada banda, dado os limites fotométricos  $m_g < 25,21$ ,  $m_r < 24,77$ ,  $m_i < 24,04$ ,  $m_z < 23,44$ ,  $m_Y < 21,70$ .

Tabela 2.7: Número de objetos por banda por intervalo de *redshift* no catálogo DES\_5000 limitado em magnitude.

<i>Redshift</i>	<i>g</i>	<i>r</i>	<i>i</i>	<i>z</i>	<i>Y</i>
$0,3 \leq z < 1,0$	75248	75247	75248	75247	73599
$1,0 \leq z < 2,0$	3368308	3368565	3365746	3091551	487831
$2,0 \leq z < 3,0$	6647248	6546033	5143020	3290513	229785
$3,0 \leq z < 4,0$	1191524	1189971	540831	268115	11027
$4,0 \leq z < 5,0$	10939	30381	12328	4727	223
$5,0 \leq z < 6,0$	389	294	1735	1758	377
$6,0 \leq z < 7,0$	0	0	1	774	200
$z \geq 7,0$	0	0	0	20	19

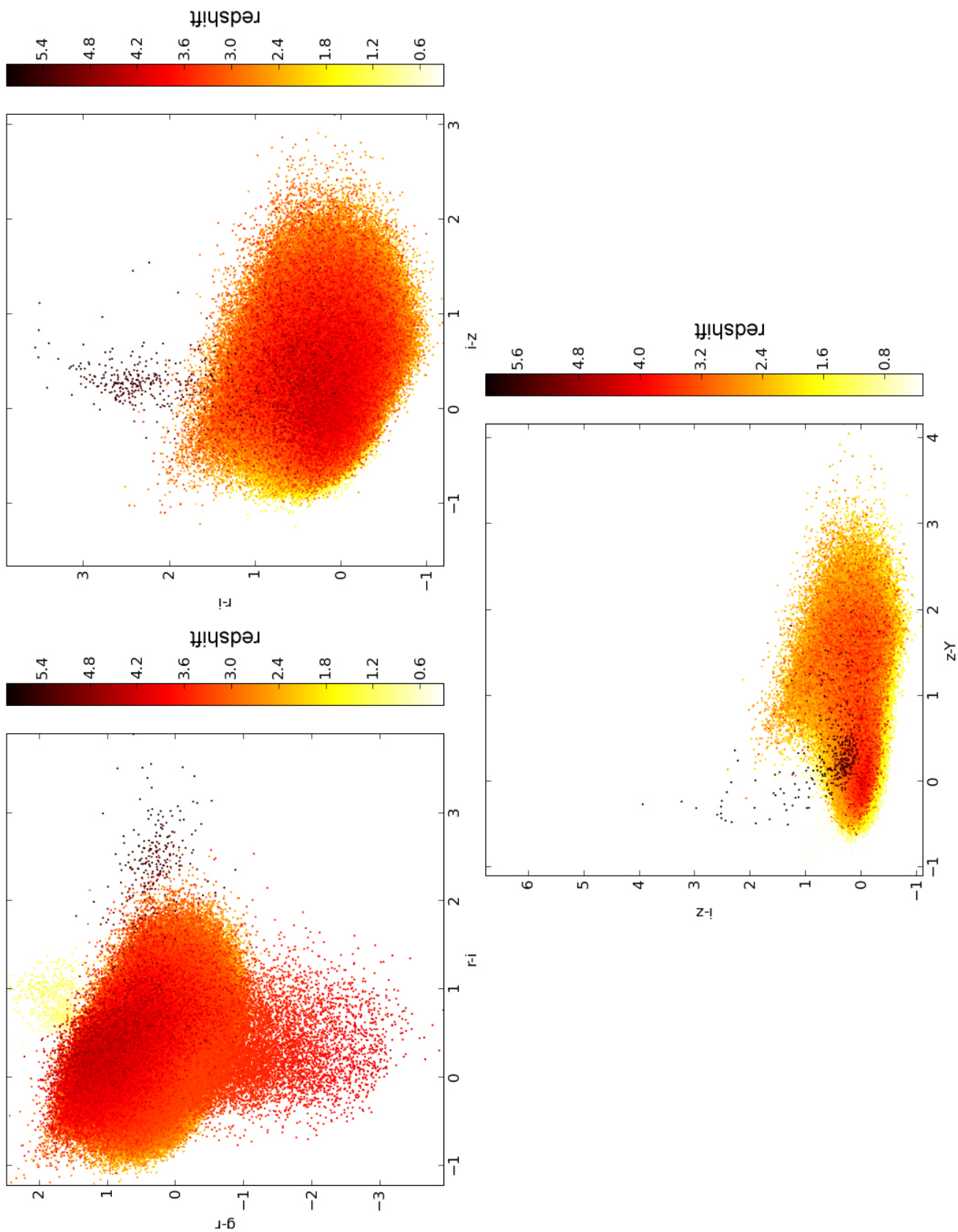


Figura 2.24: Diagramas cor-cor para os quasares do catálogo DES\_5000. A barra colorida indica o *redshift* de cada objeto.

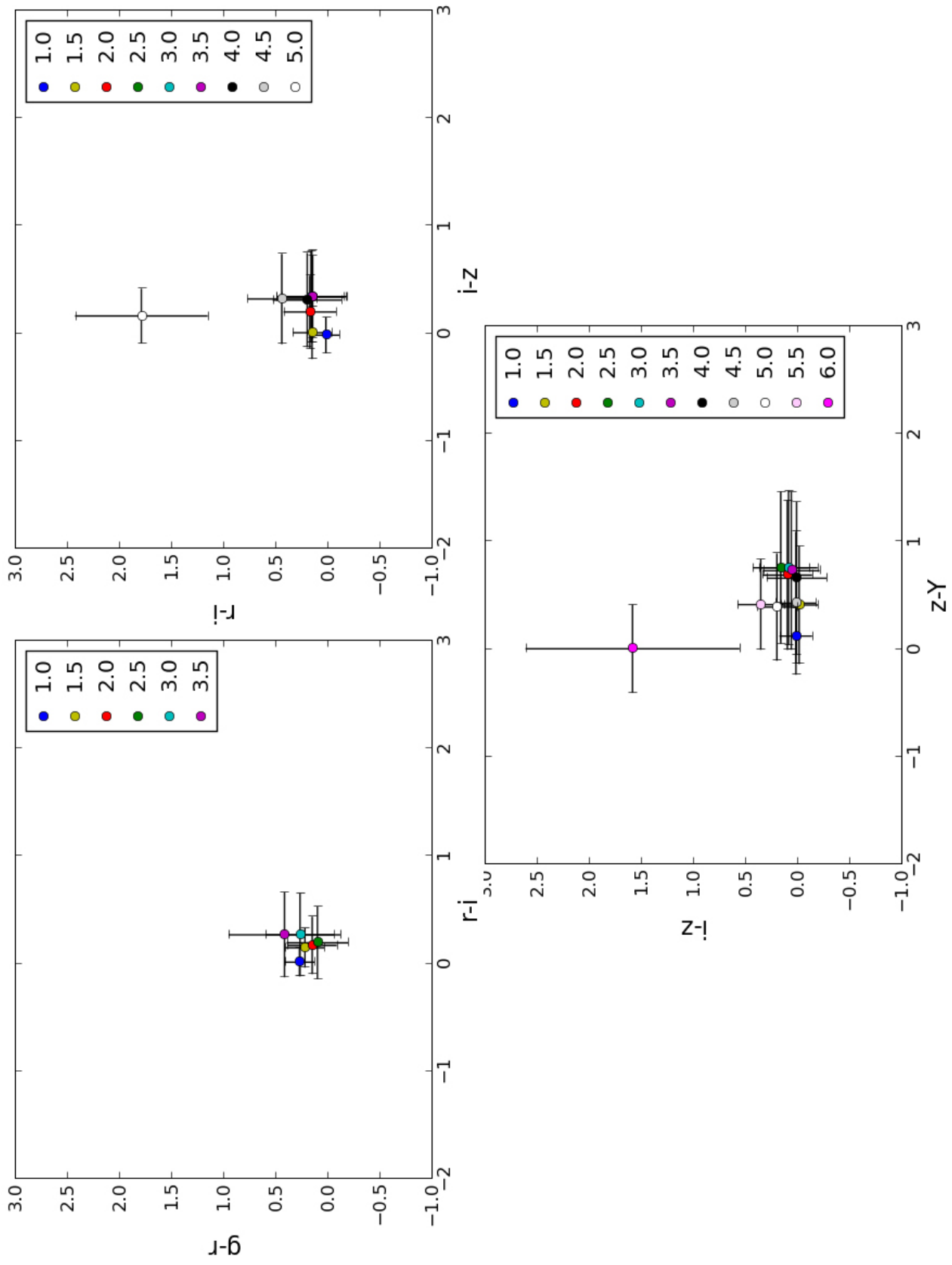


Figura 2.25: Diagramas de cor média por *redshift* dos quasares do catálogo DES\_5000.

## Capítulo 3

# Redshifts Fotométricos

O cálculo de *photo-z* surgiu nos anos 60 com Baum (1962). Com um fotômetro fotoelétrico, ele obteve informação em nove bandas para medir *redshifts* em galáxias elípticas em aglomerados de galáxias. Seu método dependia da quebra de 4000 Å como característica espectral marcante e por isso só funcionava para esse tipo morfológico de galáxias. O surgimento de levantamentos maiores e mais profundos como o *Hubble Deep Field* nos anos 90 aumentou o interesse por *photo-zs* e os grandes levantamentos atuais como o DES dependem desses resultados que se revelaram como métodos alternativos de se obter *redshifts* para amostras numerosas e profundas.

É verdade que os *photo-zs* são menos precisos que os *redshifts* espectroscópicos (*spec-zs*). De fato, em Ilbert et al. (2006) é citado que os mais acurados *photo-zs* obtidos com filtros de bandas médias (da ordem de 500 Å) ainda são por volta de trinta vezes menos acurados que *redshifts* obtidos por espectroscopia de baixa resolução. Como a obtenção de *photo-zs* depende de variações no espectro que possam ser medidas ao longo das diversas bandas quando o objeto sofre um desvio para o vermelho, quanto maior a quantidade de bandas em que se obtém a fotometria, melhor os resultados. No trabalho de Abramo et al. (2012) é utilizada uma amostra de quasares simulados em 42 filtros de banda estreita ( $\approx 100\text{\AA}$ ) para obter resultados acurados de *redshift*. Porém, acrescentar mais filtros a um levantamento aumenta bastante os custos. A solução talvez esteja na busca por sobreposições nos campos de diferentes levantamentos, o que leva a possibilidade de criação de catálogos multi-fotométricos uma vez que a combinação entre os levantamentos seja realizado.

A vantagem dos *photo-zs* está na possibilidade de medi-los de forma muito mais rápida e para muito mais objetos que os *spec-zs*. Para se obter as razões

sinal/ruído que permitem a medida dos *espec-zs*, são necessárias, relativamente à fotometria, longas integrações. Uma vez que os intervalos de comprimento de onda são muito maiores para os *photo-zs*, que, na maior parte dos casos, se baseiam em bandas médias e largas, a mesma razão sinal/ruído é obtida com muito menos tempo de exposição. Isso também possibilita obter resultados para objetos mais fracos (e mais distantes), onde a espectroscopia não é factível exceto em grandes telescópios. Além disso, até mesmo grandes espectrógrafos, como o do SDSS só conseguem obter espectros para, em média, 600 fontes por vez, reduzindo o tamanho da amostra obtida.

Os métodos para estimativas de *photo-z* encontrados na literatura podem ser divididos em ajuste de modelos e ajuste empírico. Cada tipo de ajuste apresenta vantagens e desvantagens específicas. Enquanto que métodos de ajuste de modelos tendem a apresentar maior dispersão em seus resultados quando se compara *photo-zs* a *espec-zs*, métodos de ajuste empírico, como as redes neurais, que dependem de uma amostra de treinamento que contenha *espec-zs*, tendem a apresentar maior viés, principalmente nas regiões mais próximas aos limites de *redshift* inferior e superior da amostra de treinamento.

### 3.1 Redes Neurais

O método de redes neurais para o cálculo de *photo-z* é empírico e, como tal, tem como objetivo obter uma parametrização do *redshift* em função das cores e quaisquer outros parâmetros fornecidos. Para derivar a parametrização é necessária a existência de uma amostra de treinamento, contendo valores confiáveis de *redshift*, que seja representativa dos dados a terem seus *photo-zs* calculados (amostra alvo). Nesse trabalho, utilizamos o pacote ANNz<sup>1</sup> (Collister & Lahav 2004) para calcularmos o *photo-z* utilizando redes neurais artificiais (*artificial neural networks* - *ANNs*).

O ANNz utiliza um tipo particular de ANN, que consiste em uma série de camadas dos chamados neurônios, esquematizada na figura 3.1. A estrutura sendo mostrada apresenta uma camada de neurônios com valores de entrada, uma camada

---

<sup>1</sup><http://www.homepages.ucl.ac.uk/~ucapola/annz.html>

oculta com uma quantidade menor de neurônios e uma camada de saída, com um único neurônio que no caso, carrega o *photo-z*. Cada neurônio em uma dada camada está conectado a cada um dos neurônios das camadas adjacentes e as ligações entre eles carregam um peso,  $w_{ij}$ , que juntos formam o vetor de coeficientes  $\vec{w}$  que serão otimizados. Uma *função de ativação*,  $g_j(u_j) = 1/[1 + \exp(-u_j)]$ , é definida em cada neurônio, sendo

$$u_j = \sum_i w_{ij} g_i(u_i), \quad (3.1)$$

onde a soma é feita sobre todos os neurônios  $i$  conectados ao neurônio posterior  $j$ . O neurônio de *bias* apresentado na figura 3.1 é incluído automaticamente para permitir que constantes aditivas sejam acrescentadas à função  $g_j$  (Firth et al. 2003). Para um dado vetor de entrada, o vetor de saída da rede é determinado através da progressão sequencial através das camadas, de entradas para saídas, calculando-se as ativações de cada neurônio.

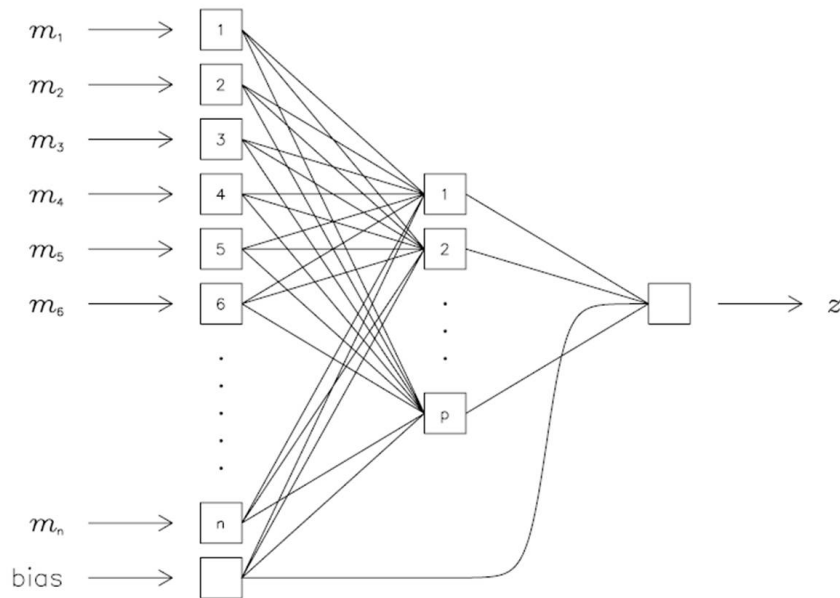


Figura 3.1: Diagrama esquemático de uma ANN com os neurônios da camada de entrada recebendo valores de magnitude em diferentes filtros e uma única camada oculta. O resultado é um neurônio na camada de saída contendo o *photo-z* (Firth et al. 2003).

Seja uma amostra de treinamento de dados de entrada  $\vec{m}$ , representando a fotometria, e  $z_{spec}$ , representando o *redshift* espectroscopicamente determinado. A

ANN é treinada através da minimização da função

$$E = \sum_k [z_{photo}(\vec{w}, \vec{m}_k) - z_{spec,k}]^2, \quad (3.2)$$

com respeito aos pesos,  $\vec{w}$ , onde  $z_{photo}(\vec{w}, \vec{m}_k)$  é a saída da rede para os dados vetores de entrada e de peso e a soma é feita sobre todos os objetos na amostra de treinamento. O ANNz utiliza o método iterativo de Quasi-Newton para realizar a minimização. Para garantir que os erros não fiquem muito grandes, um termo quadrático de custo,

$$E_w = \beta \sum_{i,j} w_{ij}^2, \quad (3.3)$$

é acrescentado à equação (3.2). O valor de  $\beta = 0.0001$  foi determinado empiricamente para otimizar a performance da ANN (Firth et al. 2003).

Após cada iteração de treinamento, a função de custo é avaliada em uma amostra separada de validação, que também contém a fotometria e os *spec-zs*. O objetivo é evitar excesso de treinamento, o que levaria a um sobre-ajuste à amostra de treinamento. Esse excesso de treinamento tende a ocorrer quando se têm muitos parâmetros sendo determinados por uma amostra de treinamento pequena. O que ocorre é um aparente aumento na eficiência do cálculo de *photo-z* pois a ANN teria memorizado os valores da amostra de treinamento. O acréscimo da amostra de validação oferece um controle à esse efeito. Dessa forma, a minimização é interrompida segundo a diminuição do erro na amostra de validação, mesmo que o erro na amostra de treinamento continue decaindo (Yèche et al 2010).

A escolha da amostra de treinamento é determinante para a obtenção de bons resultados de *photo-z*. Uma amostra ideal, seria uma subamostra representativa da amostra alvo, porém isso nem sempre é possível. Uma outra opção seria a utilização de uma amostra derivada de modelos espectrais ou de catálogos simulados, desde que a fotometria da amostra de treinamento gerada seja nos mesmos filtros e contenha as mesmas características de ruído da amostra alvo. Já o tamanho da amostra de treinamento varia de acordo com a faixa de *redshift* sendo estudada (Firth et al. 2003). O maior desafio do método de redes neurais está na dificuldade de obter resultados fora da região amostrada pelos dados de treinamento, ou seja, em extrapolar além do regime de treinamento. Isso gera resultados inverossímeis para



objetos mais fracos ou fora dos limites de *redshift* da amostra de treinamento (e também para objetos na borda dos limites), que, principalmente no caso de uma amostra real com *espec-zs*, tende a ser mais brilhante que a amostra alvo. Uma amostra de treinamento obtida através de catálogos simulados poderia contornar esse problema, porém acrescentaria outros tipos de vieses, podendo não ser representativa dos dados reais.

Para testar a eficiência do algoritmo ANNz com quasares, avaliamos seu desempenho para o catálogo QDR7. Separamos 10% da amostra aleatoriamente para formar a amostra de treinamento e outros 10% para a amostra de validação. Sendo assim, a amostra alvo era composta por cerca de 95000 objetos. Como entrada, foram requeridas 1000 iterações de treinamento, mas a convergência ocorria em torno de 900 iterações. Para testar a qualidade dos resultados, calculamos um conjunto de estatísticas que chamamos de "métricas" que inclui quatro valores: o bias, determinado como

$$\text{bias} = \langle z_{photo} - z_{espec} \rangle, \quad (3.4)$$

o desvio padrão,  $\sigma$ , o 68º percentil,  $\sigma_{68}$ , e a fração de identificações catastróficas,  $id_{catast.}$ , ou seja, aquelas em que a diferença entre o *photo-z* e o *redshift* real do objeto é superior a uma dada incerteza. Nesse trabalho, consideramos uma atribuição de *redshift* como sendo catastrófica quando

$$|z_{photo} - z_{espec}| > 0,15(1 + z_{espec}). \quad (3.5)$$

Na segunda coluna da tabela 3.1, vemos as métricas obtidas com o *photo-z* calculado pelo ANNz. Na figura 3.2 temos a comparação entre os *photo-zs* e os *espec-zs*. Há duas áreas de maior problema no cálculo de *photo-z*. A primeira, quando  $z_{espec} < 1,0$ , é decorrente da deficiência do método em determinar *redshifts* próximo às bordas da distribuição da amostra de treinamento. A segunda, quando  $z_{espec} \approx 2,2$  é provavelmente o resultado da baixa amostragem desse intervalo de *redshift* no catálogo e, conseqüentemente, na amostra de treinamento.

Uma vez que o DES não contém a banda *u* em seu conjunto de filtros, também testamos o impacto da remoção dessa banda da amostra. As métricas podem ser vistas na terceira coluna da tabela 3.1 e a comparação entre *redshifts* está na figura

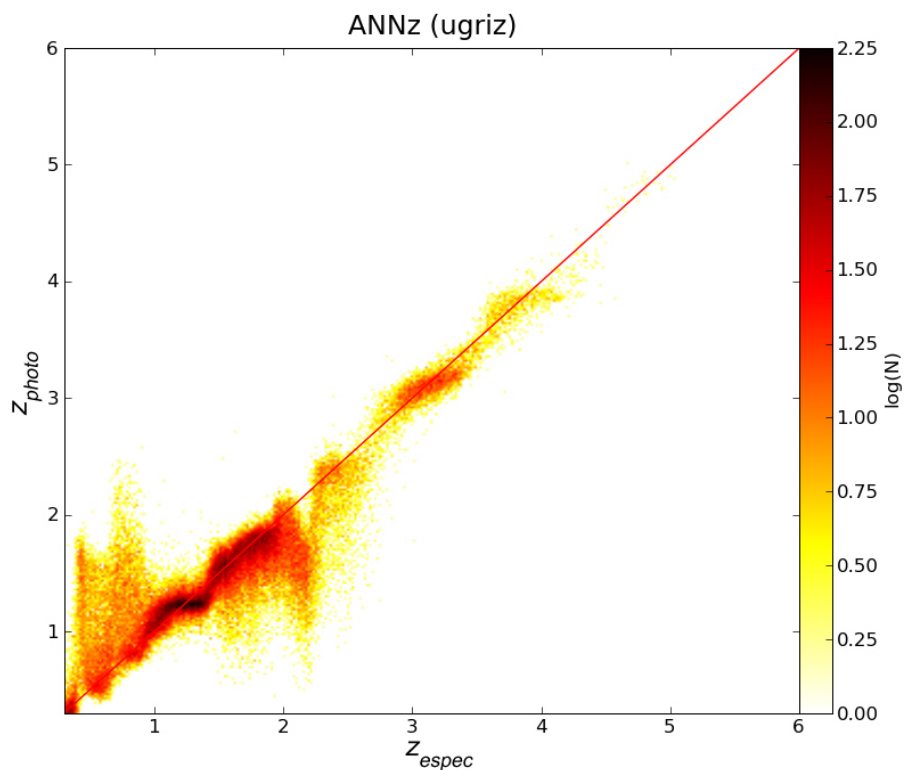


Figura 3.2: Comparação entre os  $z_{spec}$  do catálogo de quasares QDR7 e os  $z_{photo}$  calculados pelo ANNz utilizando todos os cinco filtros SDSS. A escala de cor mostra a quantidade de objetos em cada ponto.

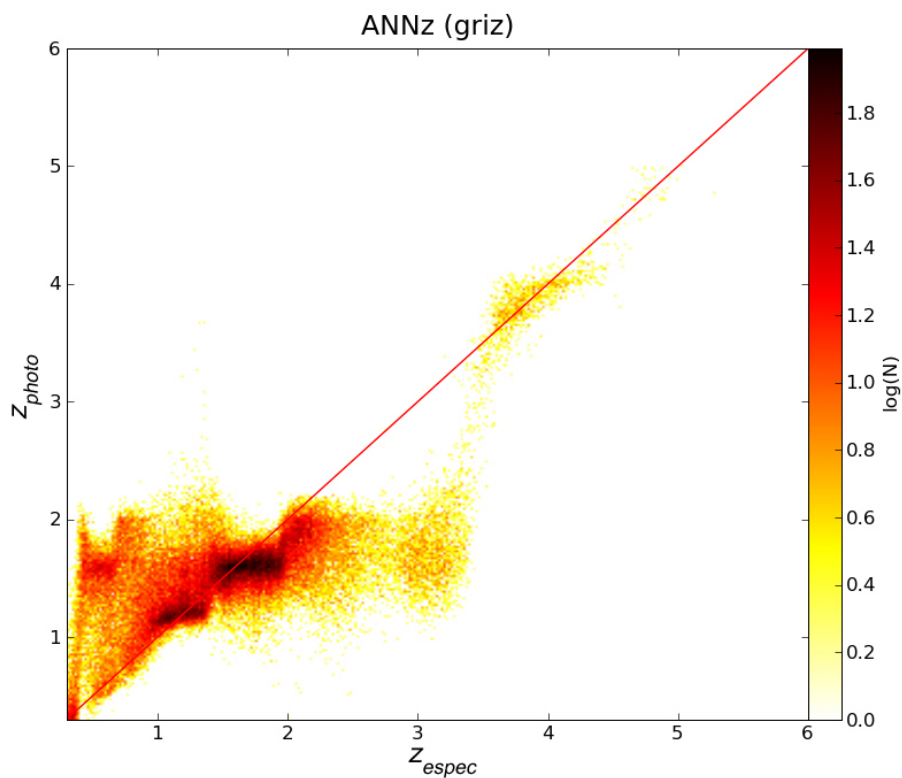


Figura 3.3: Comparação entre os  $z_{spec}$  do catálogo de quasares QDR7 e os  $z_{photo}$  calculados pelo ANNz utilizando apenas quatro filtros ( $g$ ,  $r$ ,  $i$ , e  $z$ ) SDSS. A escala de cor mostra a quantidade de objetos em cada ponto.

Tabela 3.1: Métricas do resultado do cálculo de *photo-z* do ANNz para o catálogo QDR7 com cinco e quatro bandas.

Métrica	<i>ugriz</i>	<i>griz</i>
bias	-0,004	0,008
$\sigma$	0,391	0,601
$\sigma_{68}$	0,227	0,489
Fração de $id_{catast.}$	0,216	0,392

3.3. Fica óbvia a queda na qualidade dos resultados, principalmente para objetos com  $2,0 < z_{spec} < 3,5$ . Isso é decorrente da falta da linha de  $Ly\alpha$ , que está presente na banda *u* nessa faixa de *z*. As cores nas outras bandas nessa faixa de *redshift* são muito similares, o que leva a uma degenerescência.

## 3.2 Ajuste de modelos

O método de ajuste de modelos, ao contrário do método de redes neurais, não requer uma amostra que contenha dados de *redshift* previamente calculados. Ele se baseia no uso de uma biblioteca de SEDs que podem ser obtidas de forma teórica ou através da composição de espectros reais. Sendo assim, esse método não sofre dos problemas nas bordas da distribuição de *redshift* da amostra. O *photo-z* é obtido através da comparação entre as magnitudes observadas e as magnitudes teóricas determinadas através do cálculo do fluxo esperado em cada filtro e em cada modelo de SED para vários *redshifts*. O método se baseia na detecção de características espectrais fortes que, no caso dos quasares são as linhas de emissão (principalmente  $Ly\alpha$ ,  $Ly\beta$ , C III, C IV, Mg II, Si IV,  $H\alpha$ ,  $H\beta$  e  $H\gamma$ ). Para atingir bons resultados, é indispensável a compilação de uma biblioteca de modelos representativa da amostra a ter seus *photo-zs* calculados.

Para o cálculo de *redshifts* através do ajuste de modelos, utilizamos o código LePhare<sup>2</sup> (Arnouts et al. 1999 e Ilbert et al. 2006). Através dele, cada SED é desviada para o vermelho em intervalos determinados pelo usuário (nesse trabalho utilizamos  $\delta z = 0.02$ ) e convoluída com as curvas de transmissão dos filtros (já incluindo a eficiência do instrumento). A opacidade do meio intergaláctico (Madau 1995) é considerada. O código utiliza um procedimento padrão de minimização do

<sup>2</sup><http://www.cfht.hawaii.edu/~arnouts/lephare.html>

$\chi^2$  para comparar os fluxos observados,  $F_{obs}$ , e suas incertezas,  $\sigma$ , aos fluxos dos modelos,  $F_{mod}$ :

$$\chi^2 = \sum_i \left[ \frac{F_{obs,i} - sF_{mod,i}}{\sigma_i} \right]^2, \quad (3.6)$$

onde o índice  $i$  representa cada um dos filtros presentes no levantamento estudado. Os fluxos dos modelos são normalizados para os observados através da escolha do fator  $s$  que minimiza ( $\partial\chi^2/\partial s = 0$ ) o valor de  $\chi^2$ :

$$s = \frac{\sum_i \left[ \frac{F_{obs,i}F_{mod,i}}{\sigma_i^2} \right]}{\sum_i \left[ \frac{F_{mod,i}^2}{\sigma_i^2} \right]}. \quad (3.7)$$

Idealmente, uma biblioteca de modelos deve conter todas as SEDs possíveis. Apesar disso, um aumento no número de modelos não leva necessariamente a um melhor ajuste e a valores mais acurados de *redshift*. De fato, a presença de um número muito alto de SEDs aumenta o risco de degenerescência, o que pode levar a uma quantidade mais elevada de identificações catastróficas (Salvato et al. 2009).

Para identificarmos o melhor conjunto de modelos, utilizamos uma ferramenta do LePhare que permite fixar o *redshift* de um objeto e calcula o melhor ajuste de SED para aquele  $z$ . Isso elimina o problema de degenerescência e nos fornece uma distribuição dos modelos associados aos quasares, com a possibilidade de otimizar a biblioteca utilizada. Testamos o conjunto de 44 SEDs compostas pelas bibliotecas de modelos descritas na seção 2.2 na amostra de quasares QDR7 fixando  $z$  nos respectivos valores de *spec-z* do catálogo. A distribuição final de SEDs melhor ajustadas pode ser vista na figura 3.4. Todas as SEDs, em diferentes proporções, foram ajustadas, o que nos levou a decidir por manter todo o conjunto presente ao longo de nossos testes de cálculo de *photo-z*. A distribuição de SEDs melhor ajustadas é dependente do *redshift* conforme pode ser observado na figura 3.5. Nos baseamos nesse resultado para determinarmos o sorteio de SEDs do AddQSO para gerar magnitudes a partir do conjunto das três bibliotecas.

Para testar a eficiência do LePhare com quasares, também utilizamos o catálogo QDR7. Para haver coerência entre os resultados deste código e os do ANNz, utilizamos a mesma amostra de  $\approx 95000$  quasares do QDR7 da amostra alvo da seção 3.1. Novamente, realizamos testes com e sem a banda  $u$  para verificar os efeitos sobre os resultados. A tabela 3.2 mostra as métricas para os testes e as fig-

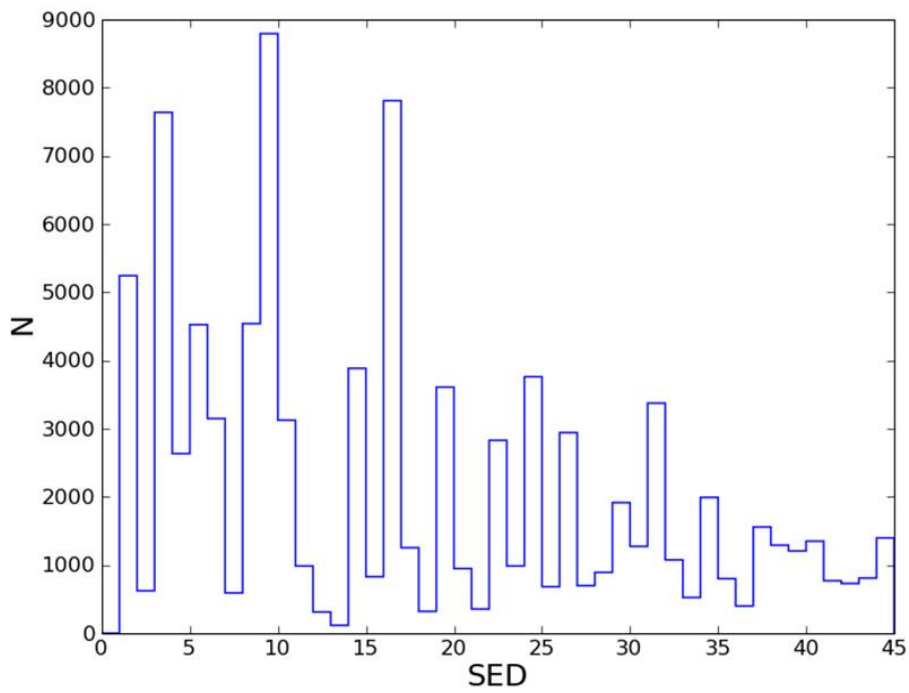


Figura 3.4: Histograma da frequência de melhor ajuste de cada SED aos quasares QDR7.

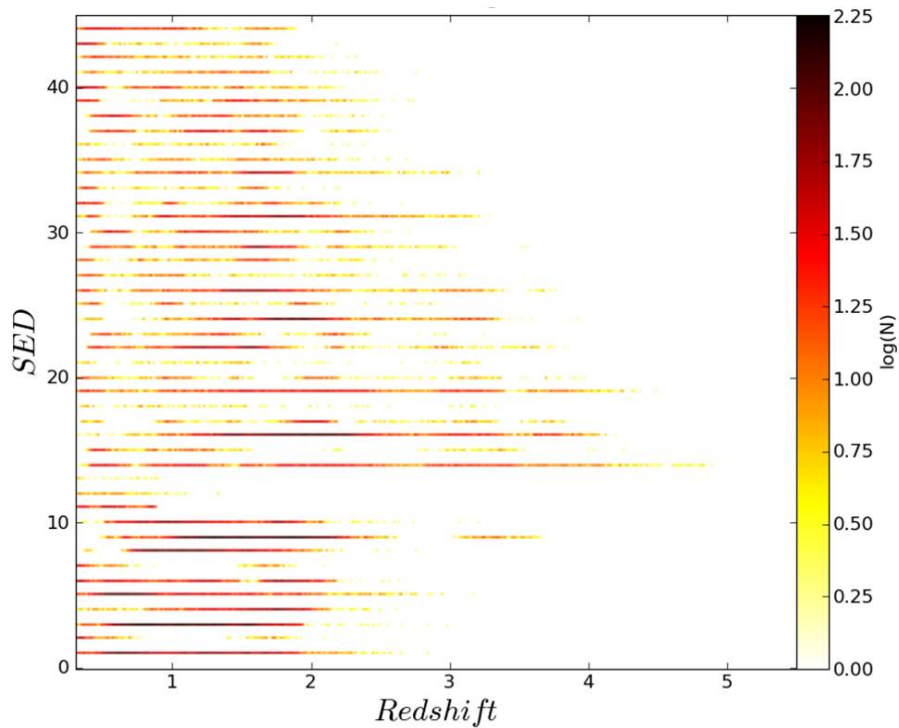


Figura 3.5: Distribuição das SEDs melhor ajustadas aos quasares do QDR7 em função do *redshift* espectroscópico do catálogo. A escala de cor mostra a quantidade de objetos em cada ponto.

Tabela 3.2: Métricas do resultado do cálculo de  $photo-z$  do LePhare para o catálogo QDR7 com cinco e quatro bandas.

Métrica	<i>ugriz</i>	<i>griz</i>
bias	-0,229	-0,170
$\sigma$	0,769	1,136
$\sigma_{68}$	0,587	1,327
Fração de $id_{catast.}$	0,357	0,531

uras 3.6 e 3.7 mostram os resultados obtidos com 5 e 4 filtros. Assim como aconteceu com o ANNz, a falta da banda  $u$  ocasiona uma piora nos resultados dos objetos com  $2,0 < z_{espec} < 3,5$ . Além disso, a remoção dessa banda faz com que o algoritmo LePhare trabalhe próximo ao seu limite de ajuste, que é a necessidade de informação em pelo menos três bandas.

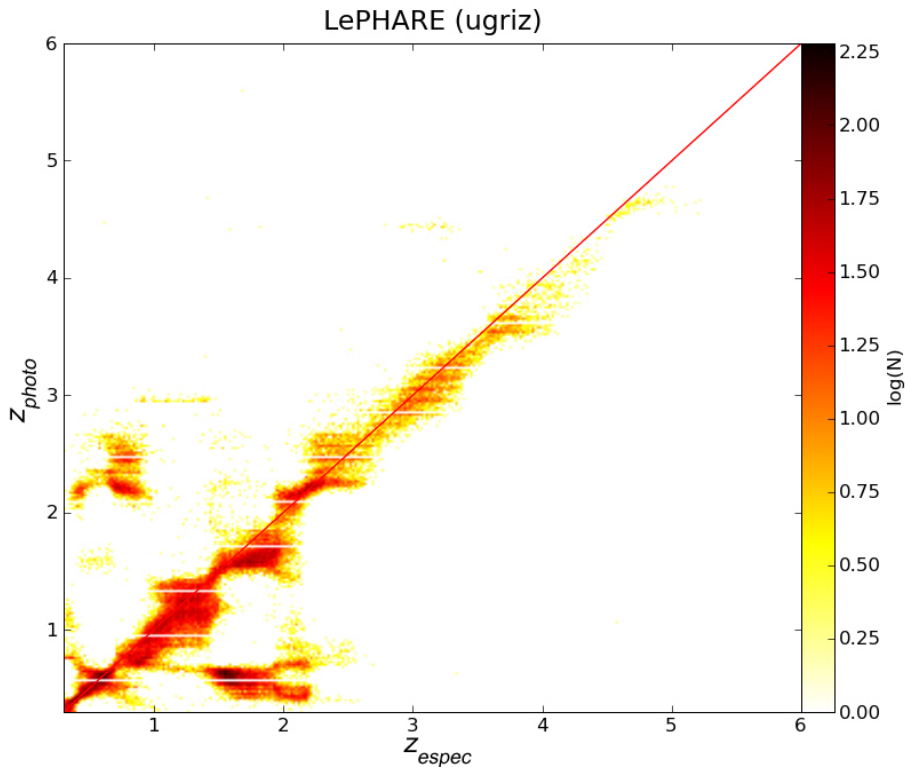


Figura 3.6: Comparação entre os  $z_{espec}$  do catálogo de quasares QDR7 e os  $z_{photo}$  calculados pelo LePhare utilizando todos os cinco filtros SDSS. A escala de cor mostra a quantidade de objetos em cada ponto.

O ajuste ruim para objetos de baixo  $z$  faz com que as métricas, mesmo no caso de todas as bandas estarem presentes, apresentem um resultado inferior ao ANNz. Isso é um efeito do método que, para quasares, apresenta identificações catastróficas naquele regime de  $z$ . A presença da linha de  $Ly\alpha$  em uma das bandas faz com que os

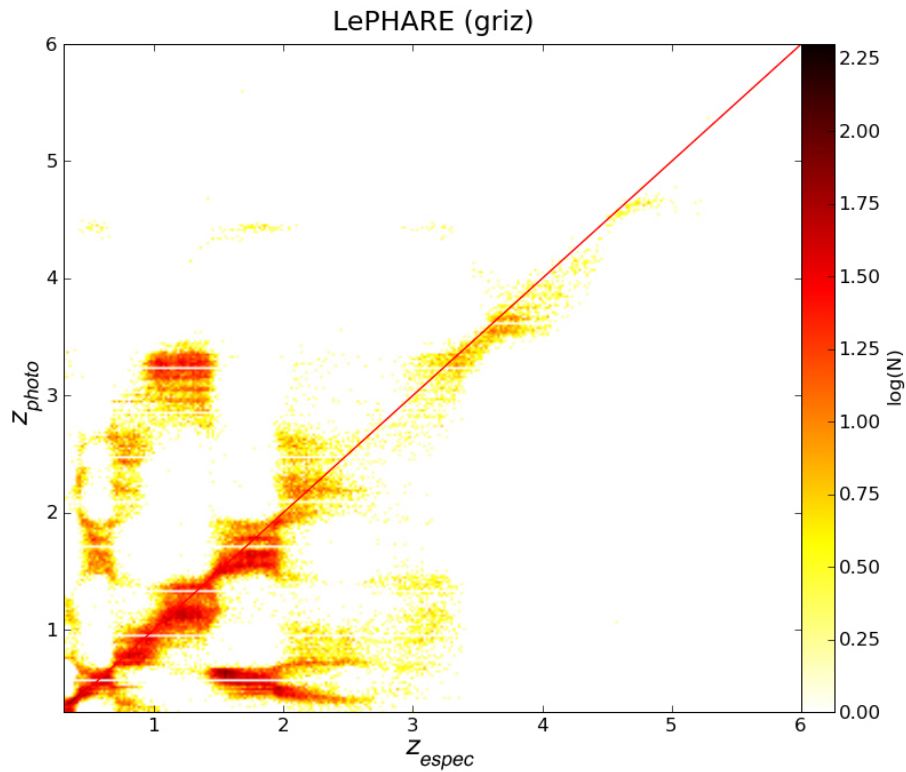


Figura 3.7: Comparação entre os  $z_{spec}$  do catálogo de quasares QDR7 e os  $z_{photo}$  calculados pelo LePhare utilizando apenas quatro filtros ( $g$ ,  $r$ ,  $i$ , e  $z$ ) SDSS. A escala de cor mostra a quantidade de objetos em cada ponto.

resultados melhorem consideravelmente para objetos de maior *redshift*, com o bias e o  $\sigma$  diminuindo, respectivamente para  $-0,111$  e  $0,446$  no caso de cinco bandas e  $z > 2,5$  e para  $-0,167$  e  $0,406$  para quatro bandas e  $z > 3,5$ . As frações de identificação catastróficas também diminuem para  $0,045$  e  $0,016$  em cada caso.

# Capítulo 4

## Classificação de Objetos

As previsões para o DES obtidas com o AddQSO mostram que na área e profundidade do levantamento são esperados mais de  $1,1 \times 10^7$  quasares. Para estudarmos esses objetos, precisaremos de uma amostra estatística dos QSOs e para tal são necessários métodos robustos capazes de selecionar amostras completas e com baixa contaminação. Além disso, para estudarmos, por exemplo, a densidade espacial, a função de luminosidade e a aglomeração, e para examinarmos a época de reionização através dos quasares, precisamos de uma amostra de QSOs de grandes *redshifts* ( $z > 6,0$ ) que precisará de *follow-up* espectroscópico para ser confirmada. Um bom método de seleção nesse intervalo de  $z$  precisa ser capaz de separar os quasares dos contaminantes mais comuns nesses *redshifts*, entre eles as estrelas de tipo tardio K e M, asteróides, galáxias de *redshift* intermediário e anãs marrons.

### 4.1 Classificação por Cor e Dropouts

A maior parte dos métodos de seleção de quasares faz a primeira classificação baseada na forma: QSOs são considerados objetos pontuais. A partir daí, os quasares são separados das estrelas através de seleções por cor. Um dos métodos mais completos de seleção de quasares é a seleção UVX baseada no excesso de radiação ultravioleta que os quasares apresentam. Esse método foi utilizado com sucesso no levantamento *2dF QSO Redshift Survey* (Croom et al. 2001), que selecionou 10000 quasares em uma área de  $\approx 290$  graus quadrados utilizando um único corte:  $18,25 < b_J < 20,85$ . Segundos os autores, o método conseguiu selecionar 90% dos QSOs entre  $0,3 < z < 2,2$ . O método, no entanto, tem um viés que leva



à seleção de objetos muito azuis e também não é eficiente em *redshifts* maiores (o objeto mais distante no catálogo de Croom et al. 2001 tem  $z \approx 3$ ). De fato, mesmo na faixa de *redshift* entre  $0,3 < z < 2,2$ , a eficiência do método fica em torno de 50% (Richards et al. 2009), ou seja, metade dos candidatos selecionados era de fato quasares. Outros métodos similares foram desenvolvidos para tentar selecionar objetos mais vermelhos, como, por exemplo, o BRX (e.g., McMahon & Irwin 1991), porém eles apresentam eficiência em torno de 35% dependendo da magnitude limite e do *redshift* (Hatziminaoglou et al. 2000).

No trabalho de Richards et al. (2009) no qual uma amostra de  $\approx 1$  milhão de quasares foi selecionada no SDSS, os autores usaram um algoritmo de classificação Bayesiana para separar quasares de estrelas em grandes latitudes galácticas. Segundo os autores, a eficiência do método é de 80%.

Para selecionar candidatos a quasares, o SDSS criou um algoritmo de seleção que busca por correspondências entre os objetos não resolvidos e o catálogo de fontes rádio FIRST e separa objetos de cores diferentes das estrelas no espaço de cor do levantamento (Richards et al. 2002). Há dois espaços de cores tridimensionais para candidatos de pequenos (baseado nas cores  $u - g$ ,  $g - r$  e  $r - i$ ) e grandes (baseado em  $g - r$ ,  $r - i$  e  $i - z$ ) *redshifts*. Os candidatos são selecionados se estão a mais de  $4\sigma$  do ponto mais próximo do locus estelar. O algoritmo, quando seleciona objetos no espaço de cor que envolve as bandas  $u$ ,  $g$ ,  $r$ , e  $i$ , também considera candidatos extensos para incluir na amostra os quasares de pequeno *redshift* cuja hospedeira é visível, além de outros tipos de AGN. Para não selecionar galáxias extensas, cortes em cor são realizados. O locus das galáxias não foi incluído. Richards et al. (2002) concluem que o algoritmo de seleção produz uma amostra completa acima de 90% com uma eficiência de 65%.

Os exemplos citados selecionam quasares relativamente brilhantes em *redshifts* inferiores a 6,0. Uma amostra mais profunda e que contenha objetos com *redshifts* maiores irá sofrer com a contaminação não só de estrelas, mas também de galáxias. O algoritmo do SDSS considera a população de galáxias extensas como contaminante na amostra de candidatos de pequeno *redshift*, mas não leva em conta a contaminação desses objetos quando eles demonstram morfologia similar aos quasares. Um algoritmo similar pode ser utilizado pelo DES, mas é provável que a falta da

banda  $u$  prejudique a seleção de candidatos de pequenos *redshifts* e que os limites de magnitude mais profundos venham atrelados a uma queda na eficiência devido à contaminação por galáxias.

Para selecionar candidatos a quasares de *redshifts* superiores a 6,0, o SDSS utilizou a técnica de seleção por *i-dropouts* (e.g., Fan et al. 2003). O critério de classificação de *i-dropout* é a falta de detecção nas bandas  $u$ ,  $g$  e  $r$ ,  $m_z < 20,2$ ,  $\sigma(z) < 0,1$ ,  $i-z > 2,2$  e  $z-J < 1,5+0,35(i-z-2,2)$ , onde a fotometria na banda  $J$  é obtida através da correspondência das fontes com objetos de outros levantamentos ou através de uma medida independente. A motivação por trás desse critério de seleção é o fato de que para *redshifts* superiores a 5,7 a linha de  $\text{Ly}\alpha$  começa a aparecer em comprimentos de onda superiores aos do filtro  $i$ , o que permite separar objetos de grandes *redshifts* da maioria dos objetos de aparência estelar. No entanto, a maior parte dos candidatos selecionados por esse critério são falsas detecções na banda  $z$  causadas por raios cósmicos, e estrelas de tipo M, L e T. A técnica de seleção de *dropouts* pode ser estendida para objetos com *redshifts* superiores a 7, como foi feito para galáxias buscando por *z-dropouts* (e.g., Ono et al. 2012). Para o DES, as informanções nas bandas  $J$ ,  $H$  e  $K_s$  do VHS serão indispensáveis para selecionar quasares de grandes *redshifts*.

## 4.2 Seleção por Ajuste de Modelos

O método de ajuste de modelos descrito na seção 3.2 pode ser aplicado como uma forma de classificação de objetos. Para isso, precisamos testar o ajuste de SEDs de galáxias, estrelas e quasares. O LePhare permite que isso seja feito em uma única rodada. O usuário fornece as bibliotecas de SEDs de cada tipo de objeto e o algoritmo calcula o melhor ajuste para cada tipo. É possível então atribuir uma "classe" como sendo a do tipo ajustado com menor valor de  $\chi^2$ .

A biblioteca de modelos de quasares escolhida para nossos testes é composta pelo conjunto de 44 SEDs descritas na seção 2.2. Já para a biblioteca de estrelas, optamos por usar um conjunto que inclui as 131 SEDs de Pickles (1998), 19 SEDs com  $\log g = 4.00$ , onde  $g$  é a gravidade superficial em  $\text{cm s}^{-2}$ , e temperaturas variando entre 3500 K a 8000 K, e quatro espectros de anãs brancas obtidos por Bohlin et al.

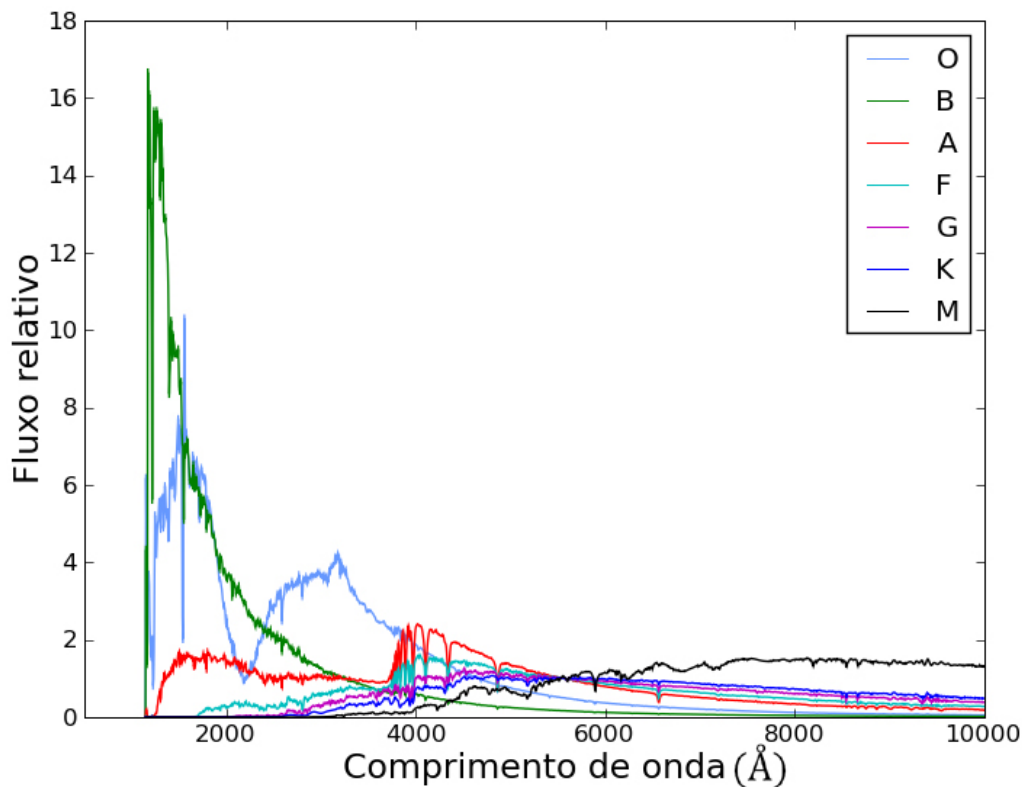


Figura 4.1: Subconjunto de SEDs de estrelas presentes no LePhare. Estão representadas 8 SEDs do conjunto de Pickles (1998).

(1995). A biblioteca de Pickles contém espectros calibrados em fluxo que incluem todos os tipos espectrais e classes de luminosidade em abundância solar, além de anãs F, G e K e gigantes G e K fracas em metal. Na figura 4.1 apresentamos alguns dos espectros dessa biblioteca.

Por último, a biblioteca de modelos de galáxias contém seis modelos. As galáxias espirais, irregulares, Sbc e Scd são representadas cada uma por um espectro obtidos de Coleman et al (1980) e extrapoladas para comprimentos de onda do ultravioleta e do infravermelho. Os dois modelos restantes são de galáxias com surto de formação estelar provenientes do atlas de Kinney (Calzetti et al. 1994). Os seis modelos podem ser vistos na figura 4.2.

Para validar o ajuste de modelos como forma de classificação, utilizamos, assim como no cálculo de *photo-z*, o catálogo de quasares QDR7. Novamente, testamos nossos resultados na presença da banda *u* e a removemos para avaliarmos o impacto. Na tabela 4.1 mostramos a porcentagem de classificação em cada classe segundo o critério de menor  $\chi^2$ . A falta da banda *u* afeta o cálculo do *photo-z*, ou seja, a qualidade do ajuste de SED. Sendo assim, é de se esperar uma piora na classificação.

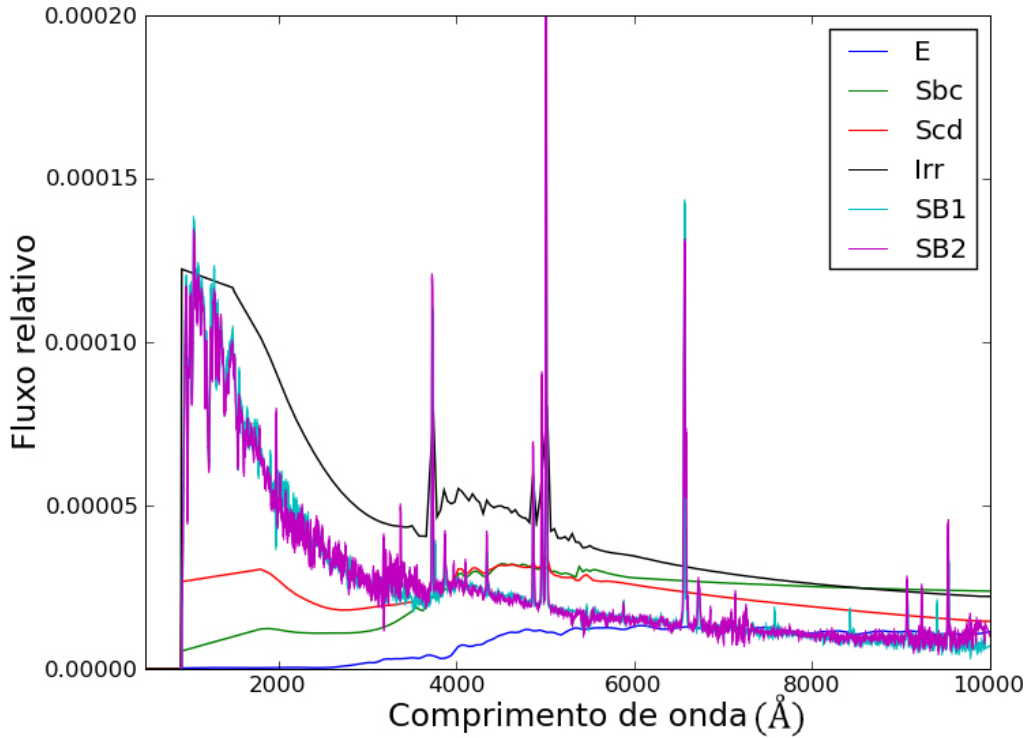


Figura 4.2: Conjunto de SEDs de galáxias presentes no LePhare. Todas as SEDs da biblioteca estão representadas.

Tabela 4.1: Porcentagem de classificação dos quasares do catálogo QDR7 segundo o menor  $\chi^2$  obtido com o LePhare.

Classe	<i>ugriz</i>	<i>griz</i>
estrela	0,39%	2,61%
galáxia	7,81%	8,75%
quasar	91,80%	88,64%

Pela tabela, fica claro que a maior parte dos quasares classificados erroneamente é confundido com galáxias.

Além da taxa de classificação, é interessante avaliar que tipo de SED de galáxia ou estrela é ajustada aos quasares e leva a uma classificação errada. Nas figuras 4.3 e 4.4 apresentamos, respectivamente, histogramas com a frequência de ajuste de cada SED estelar ou de galáxia com e sem a inclusão da banda *u*. As SEDs de estrelas são numeradas de acordo com o que pode ser visto na tabela 4.2. As SEDs de galáxias, por sua vez, são numeradas conforme é mostrado na tabela 4.3.

Fica claro que as SEDs de estrela que melhor se ajustam aos quasares diferem na presença e na ausência da banda *u*. Em ambos os casos, as SEDs de tipos estelares mais quentes e de anãs quase não são ajustadas. Porém, no caso da ausência da banda *u* há uma preferência por ajustes de SEDs de estrelas de tipo F e G e de SEDs

Tabela 4.2: Identificação numérica dos tipos de SEDs de estrelas que podem ser ajustadas pelo algoritmo LePhare para classificar objetos.

1 – 3	tipo O
4 – 22	tipo B
23 – 35	tipo A
36 – 55	tipo F
56 – 78	tipo G
79 – 110	tipo K
111 – 131	tipo M
132 – 135	anãs brancas
136 – 154	$\log g = 4.00$ e $3500 \text{ K} \leq T \leq 8000 \text{ K}$

Tabela 4.3: Identificação numérica dos tipos de SEDs de galáxias que podem ser ajustadas pelo algoritmo LePhare para classificar objetos.

1	galáxia de tipo espiral
2	galáxia de tipo Sbc
3	galáxia de tipo Scb
4	galáxia irregular
5 – 6	galáxias com surto de formação estelar

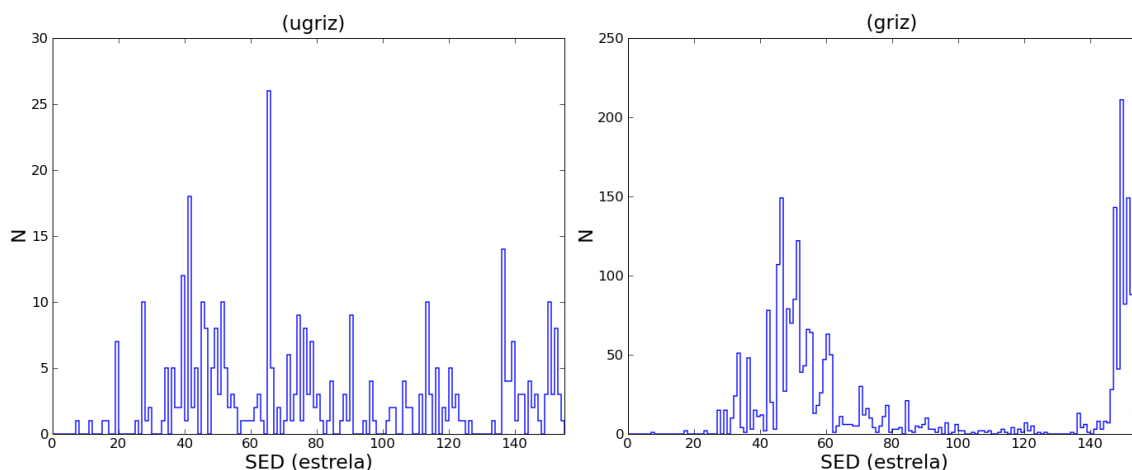


Figura 4.3: Histogramas da frequência do ajuste das SEDs de estrela aos quasares do QDR7 classificados erroneamente como estrelas pelo LePhare. No painel da esquerda está o resultado obtido na presença da banda  $u$ . No painel da direita está o resultado sem a banda  $u$ .

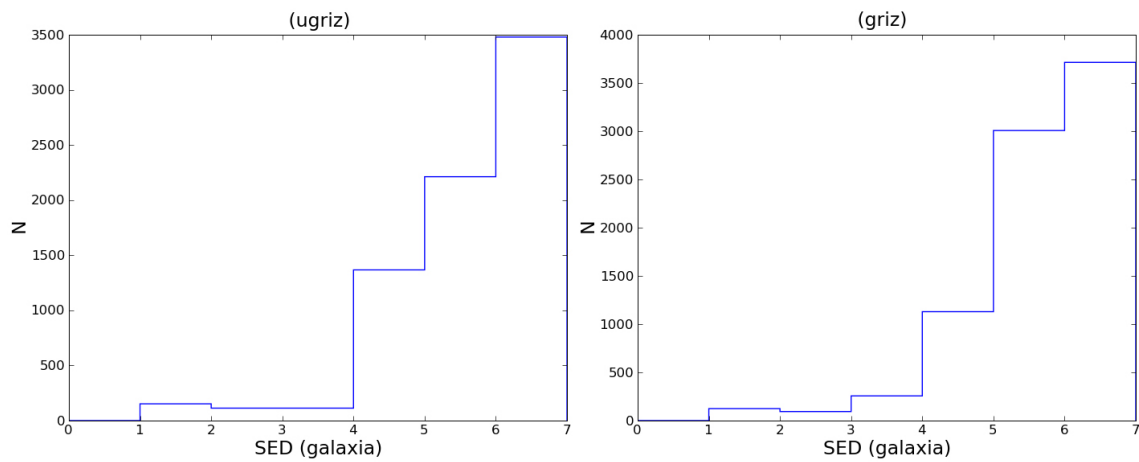


Figura 4.4: Histogramas da frequência do ajuste das SEDs de galáxia aos quasares do QDR7 classificados erroneamente como galáxias pelo LePhare. No painel da esquerda está o resultado obtido na presença da banda  $u$ . No painel da direita está o resultado sem a banda  $u$ .

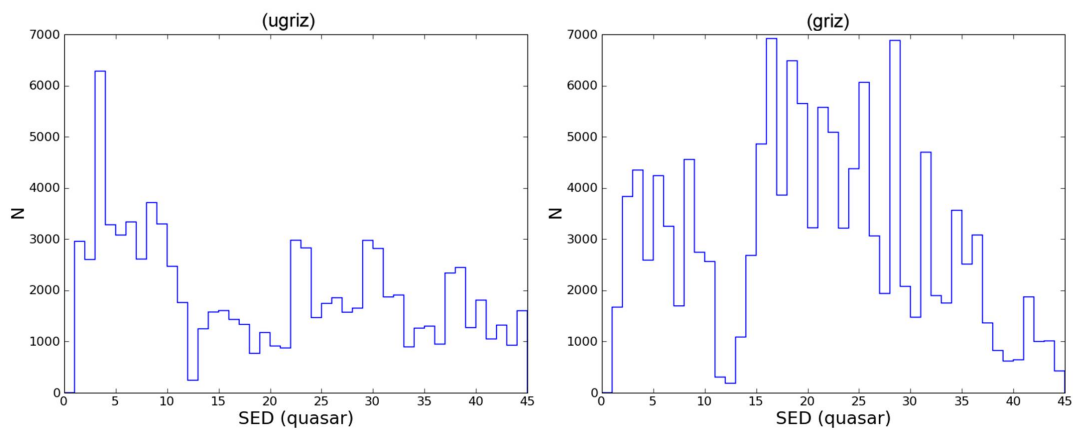


Figura 4.5: Histogramas da frequência do ajuste das SEDs de quasar aos objetos do QDR7 classificados corretamente como QSOs pelo LePhare. No painel da esquerda está o resultado obtido na presença da banda  $u$ . No painel da direita está o resultado sem a banda  $u$ .

de estrelas com gravidade  $\log g = 4.00$  e  $T > 6500$  K. Já entre as SEDs de galáxias, as de surto de formação estelar são melhor ajustadas independente da presença da banda  $u$  nos testes.

Já a distribuição de SEDs de quasar melhor ajustadas aos QSOs do QDR7 corretamente classificados pelo LePhare pode ser vista na figura 4.5 na presença e na ausência da banda  $u$ . No primeiro caso, há uma preferência pelas SEDs de Budavári et al. (2001), representadas pelos números 1 a 8. No caso sem a banda  $u$  esse perfil muda, havendo maior preferência pelas SEDs com números entre 15 e 37, que compõe a biblioteca de modelos sintéticos. Fica claro que, na falta da banda  $u$ , a distribuição de SEDs obtidas com o melhor ajuste sem fixar o *redshift* difere bastante daquela obtida na figura 3.4. Isso é um indicativo da degenerescência do método, ou seja, as cores do objeto podem ser bem ajustadas por uma SED com um determinado *photo-z* e por outra SED diferente com um outro *redshift*. O mais provável é que essa degenerescência ocorra principalmente quando a linha de Ly $\alpha$  ocupa comprimentos de onda correspondentes ao filtro  $u$ .

# Capítulo 5

## Aplicação das ferramentas de análise ao DES

Nos capítulos anteriores, explicamos o funcionamento das ferramentas de análise e validamos seus resultados utilizando o catálogo de quasares espectroscopicamente confirmados do DR7 do SDSS. Agora, iremos aplicar essas ferramentas aos quasares simulados pelo AddQSO. Dessa forma, podemos medir a qualidade dos resultados esperados com o DES e determinar quais são as melhores ferramentas para atingir nossos objetivos, seja ele a classificação de objetos ou a determinação do *photo-z*.

### 5.1 Os catálogos

Diferentes catálogos foram gerados pelo AddQSO. Eles variam em área, intervalo de *redshift*, magnitudes limite entre outros parâmetros. Nomeados de acordo com os filtros utilizados e a área do catálogo, listamos suas principais características na tabela 5.1. A área é fornecida em graus quadrados, a QLF é definida como 1 para Richards et al. (2006), 2 para Willott et al. (2010) e 3 para a junção de ambas as QLFs feita no *redshift*  $z_{join}$ . Os parâmetros  $z_{min}$ ,  $z_{max}$ ,  $M_{i,min}$  e  $M_{i,max}$  são referentes aos limites fornecidos para a integração da QLF. O número de quasares se refere ao total de objetos gerados, sem o emprego de cortes em magnitude. Todas as magnitudes aparentes nas simulações do AddQSO foram geradas a partir do conjunto de 44 SEDs. Na tabela 5.1 também listamos as características dos catálogos de simulações de estrelas e galáxias do DC6B. Esses catálogos auxiliares têm o prefixo "DC" para identificá-los. Por último, também caracterizamos os catálogos



QDR7 e QR09 usados para testar as ferramentas de análise e validar os algoritmos do AddQSO.

Tabela 5.1: Características gerais da geração dos catálogos simulados utilizados nos testes das ferramentas de análise e dos catálogos de quasares da literatura (QDR7 e QR09).

Nome	Área <sup>†</sup>	Filtros	Nº de QSOs <sup>††</sup>	QLF <sup>‡</sup>	$M_{i,min}$	$M_{i,max}$	$z_{min}$	$z_{max}$	$z_{join}$ <sup>‡‡</sup>
DES_5000	5000	DES	11928502	3	-31,0	-22,9	0,3	8,0	5,25
SDSS_8417	8417	SDSS	16603122	1	-31,0	-22,9	0,3	5,5	—
DES_66	66,83	DES	159879	3	-22,9	-23,0	0,3	8,0	5,25
DC6B_S	66,83	DES	1092476	—	—	—	—	—	—
DC6B_G	66,83	DES	2364152	—	—	—	—	—	—
QDR7	9380	SDSS	105783	—	—	—	0,065	5,46	—
QR09	8417	SDSS	1172157	—	—	—	—	—	—

<sup>†</sup>em graus quadrados

<sup>††</sup>sem cortes em magnitude

<sup>‡</sup> 1 = Richard et al. (2006); 2 = Willott et al. (2010); 3 = ambas as QLFs

<sup>‡‡</sup>*redshift* onde as duas QLFs foram unidas

## 5.2 Classificação

Para analisar os resultados do processo de classificação, precisamos, além de um catálogo de quasares, catálogos de estrelas e galáxias com os quais podemos medir a taxa de contaminação. Para nossos testes, utilizamos os dados da simulação DC6B da Colaboração DES em uma área total de 66,83 graus quadrados. Para criar os catálogos de estrelas e galáxias selecionamos apenas os objetos classificados corretamente pela tabela verdade. Eles compõem os catálogos DC6B\_S, com 1092476 estrelas e DC6B\_G, com 2364152 galáxias, sem cortes em magnitude.

Para o catálogo de quasares, DES\_66, simulamos esses objetos em uma área de 66,83 graus quadrados, similar aos catálogos de estrelas e galáxias. A simulação foi realizada pela integração de ambas as QLFs unidas em  $z_{join} = 5,25$ . O intervalo de redshift para a integração foi de  $0,3 < z < 8,0$  e o intervalo de magnitude absoluta na banda  $i$  foi de  $-31,0 < M_i < -22,9$ . As magnitudes aparentes foram geradas utilizando o conjunto de 44 SEDs. Erros fotométricos foram acrescentados às magnitudes. O total de objetos gerados é 159879. Assim como em DC6B\_S e DC6B\_G, esse valor representa o número de quasares sem cortes em magnitude.

### 5.2.1 Seleção por cor e *dropouts*

Antes de testarmos o ajuste de modelos como forma de classificação de objetos, avaliamos a distribuição em diagramas cor-cor para cada um de nossos catálogos. Como todos os três catálogos são simulados, podemos examinar o impacto do acréscimo de erros fotométricos às magnitudes presentes na tabela verdade. Os resultados podem ser vistos nas figuras 5.1 para as cores  $g - r \times r - i$ , 5.2 para as cores  $r - i \times i - z$  e 5.3 para as cores  $i - z \times z - Y$ .

Não é possível separar completamente os quasares das estrelas e galáxias baseado em suas cores disponíveis pelo levantamento DES. Se selecionarmos as regiões dos diagramas em que não há contaminação por outros objetos, conseguimos separar, no máximo 2,5% dos quasares no diagrama  $i - z \times z - Y$ . Se considerarmos apenas a contaminação por estrelas, é possível selecionar  $\approx 27,3\%$  dos quasares no diagrama  $g - r \times r - i$ , o que ainda é inferior à eficiência dos métodos de seleção presentes na literatura.

Como o catálogo DES\_66 contém poucos objetos, principalmente em grandes *redshifts* devido a sua pequena área de cobertura, para termos uma estatística da quantidade de *dropouts* por banda utilizamos o catálogo DES\_5000, que foi descrito na seção 2.6. Para um dado filtro  $x$ , consideramos um  $x$ -*dropout* o quasar que apresenta  $m_x$  mais fraca que a magnitude limite do levantamento, e magnitude na banda seguinte,  $m_y$ , dentro dos limites no levantamento. Não são impostos limites sobre as bandas que englobam comprimentos de onda maiores que os do filtro  $y$ . O resultado pode ser visto na tabela 5.2 que mostra o número total de *dropouts* esperados por banda e a porcentagem desses objetos em relação à amostra total de 11668480 quasares que apresentam ao menos uma banda dentro dos limites esperados pelo DES. Os efeitos da floresta de  $\text{Ly}\alpha$  são os responsáveis pelo grande número de  $g$ -*dropouts*. Como não temos filtros em comprimentos de onda maiores que os da banda  $Y$ , não calculamos  $Y$ -*dropouts*.

Tabela 5.2: Contagem de *dropouts* por banda presentes no catálogo DES\_5000.

	número de <i>dropouts</i>	fração do total de objetos
$g$ - <i>dropouts</i>	324728	2,78%
$r$ - <i>dropouts</i>	37648	0,323%
$i$ - <i>dropouts</i>	12524	0,107%
$z$ - <i>dropouts</i>	95	$8,14 \times 10^{-4}\%$

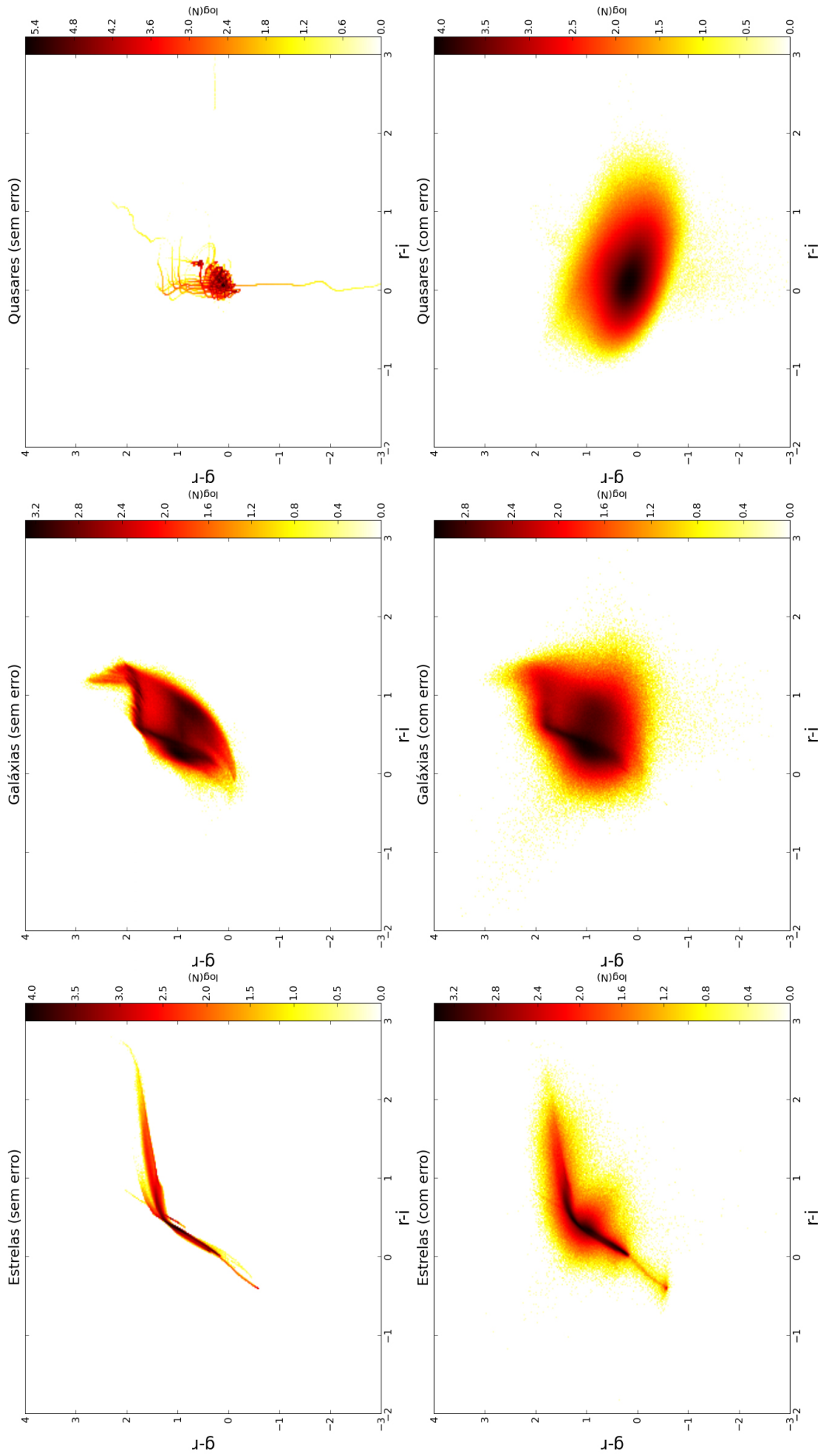


Figura 5.1: Diagramas cor-cor de  $g - r \times r - i$  para os objetos dos catálogos DC6B\_S\_66 de estrelas, DC6B\_G\_66 de galáxias e DES\_66 de quasares. Nos painéis superiores, as cores foram obtidas com base nas magnitudes das tabelas verdade sem acréscimo de erros fotométricos. Nos painéis inferiores as cores são calculadas com base em magnitudes com erros fotométricos.

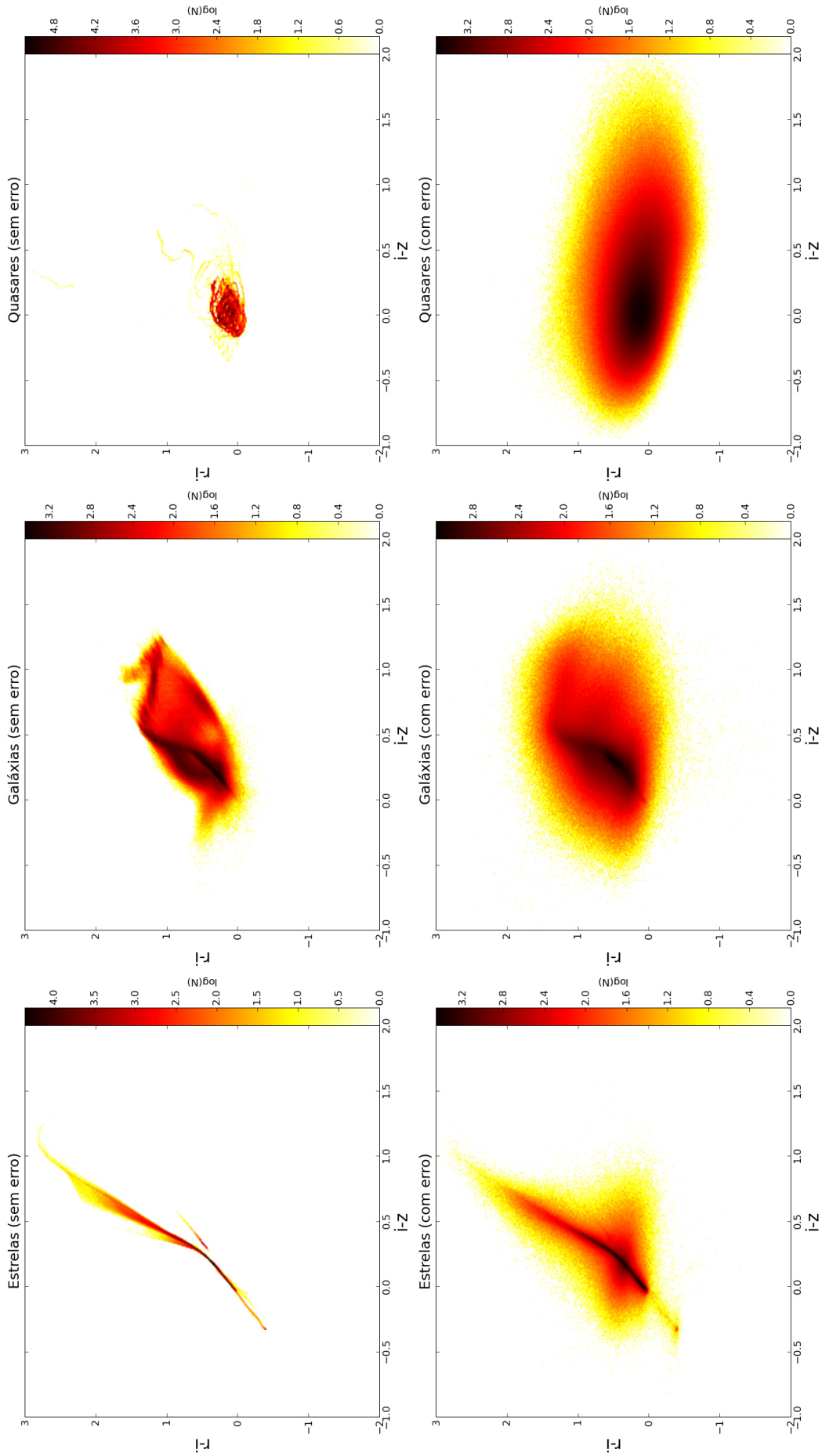


Figura 5.2: O mesmo que na figura 5.1 para  $r - i \times i - z$ .

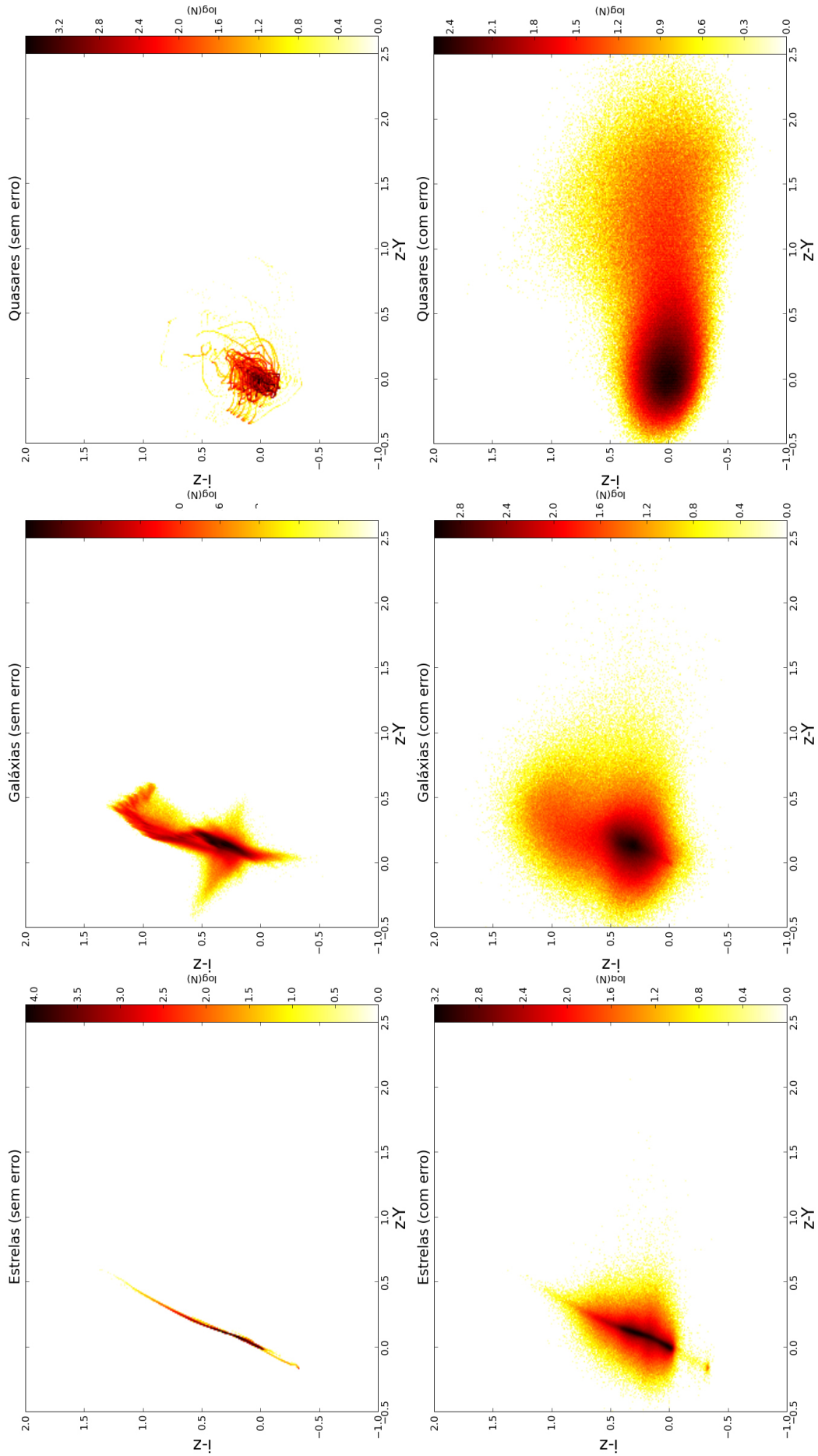


Figura 5.3: O mesmo que na figura 5.1 para  $i = z \times z - Y$ .

Conforme foi mencionado na seção 4.1, para selecionar candidatos a quasares com *redshifts* superiores a 6,0 e 7,0, buscamos, respectivamente, por *i* e *z-dropouts*. Se aplicarmos as porcentagens da tabela 5.2 ao catálogo DES\_66, podemos estimar a contaminação por estrelas e galáxias, observando a incidência de *dropouts* nos catálogos DC6B\_S e DC6B\_G. O resultado é mostrado na tabela 5.3. As estrelas e galáxias são mais numerosas que os quasares e, como esperado, apresentam uma maior quantidade de *dropouts* numa mesma área. Como as simulações não contém fotometria no infravermelho, não podemos aplicar os critérios de seleção em  $z - J$  como foi feito para determinar *i-dropouts* no SDSS, então os resultados da tabela 5.3 são apenas ilustrativos do problema de contaminação por outros tipos de objetos no critério de seleção por *dropouts*. No entanto, no trabalho de Fan et al. (2003) em que a busca por *i-dropouts* levou a descoberta de três quasares com *redshifts* superiores a 6, a eficiência do critério de seleção por cor era de apenas 2%, o que é similar ao que vemos na tabela 5.3.

Tabela 5.3: Contagem de quasares, estrelas e galáxias *dropouts* nas bandas *i* e *z*.

	<i>i-dropouts</i>	<i>z-dropouts</i>
DES_66	171	1
DC6B_S	1504	69
DC6B_G	7293	113

## 5.2.2 Classificação por ajuste de modelos

A qualidade da classificação por ajuste de SEDs é medida por dois fatores: a taxa de classificação correta dos quasares e a taxa de contaminação por estrelas e galáxias, ou, analogamente, a completeza e a pureza de um catálogo de candidatos a quasares selecionados por esse método. Sejam *S*, *G* e *Q* representativos das três classes (estrelas, galáxias e quasares) que um objeto pode assumir. O número de objetos de classe *X* corretamente classificado será identificado por  $X_X$  e o número de objetos classificados incorretamente como pertencentes a outras classes será identificado por  $Y_X$  e  $Z_X$ . Na tabela 5.4 definimos as variáveis considerando as três classes possíveis.

Tabela 5.4: Classificações possíveis para cada tipo de objeto.

	Estrelas	Galáxias	Quasares
Classificados como estrelas	$S_S$	$S_G$	$S_Q$
Classificados como galáxias	$G_S$	$G_G$	$G_Q$
Classificados como quasares	$Q_S$	$Q_G$	$Q_Q$

Para uma amostra de quasares, a completudeza é definida como:

$$c^Q = \frac{Q_Q}{S_Q + G_Q + Q_Q}. \quad (5.1)$$

A pureza, por sua vez, é análoga a eficiência e é definida conforme:

$$p^Q = \frac{Q_Q}{Q_S + Q_G + Q_Q}. \quad (5.2)$$

Sendo assim, a taxa de contaminação é dada por:

$$f = 1 - p^Q. \quad (5.3)$$

Utilizamos o LePhare para ajustar as SEDs descritas na seção 4.2 aos objetos.

No caso do catálogo DES\_66, a taxa de classificação pode ser vista na tabela 5.5.

Tabela 5.5: Classificação dos quasares do catálogo DES\_66 segundo a menor  $\chi^2$  obtida com o LePhare.

	número de objetos	fração da amostra
$S_Q$	5736	4,01%
$G_Q$	5666	3,97%
$Q_Q$	131485	92,02%

O resultado é similar àquele obtido para os quasares do DR7 do SDSS na seção 4.2 na ausência da banda  $u$ . A maior parte dos quasares é corretamente classificada porém, com o catálogo DES\_66 as perdas de objetos classificados incorretamente são comparáveis nas classes de estrelas e galáxias. A completudeza é de  $c^Q = 0,92$ . Esse valor é comparável com os melhores métodos de seleção presentes na literatura.

A pureza, no entanto, não é satisfatória. As classificações das estrelas e das galáxias dos catálogos DC6B\_S e DC6B\_G, respectivamente, podem ser vistas na tabela 5.6. Como pode ser observado, o método não classifica bem esses tipos de objetos, com a completudeza da amostra de estrelas sendo de apenas  $c^S = 0,26$  e a

da amostra de galáxias de  $c^G = 0,39$ . A classificação ruim reflete na obtenção de uma amostra de quasares selecionados por ajuste de modelos. Baseado no resultado para os catálogos DC6B\_S e DC6B\_G, uma amostra de quasares selecionados em uma área de 66,83 graus quadrados teria pureza  $p^Q = 0,06$ , ou seja, uma taxa de contaminação de quase 0,94. Se removermos a contaminação por galáxias argumentando que esses objetos podem ser eliminados por critérios de morfologia, a pureza da amostra de quasares aumenta para  $p^Q = 0,2$ , ou uma contaminação estelar de mais de 0,8, o que ainda é bem aquém dos métodos de classificação de objetos da literatura. O *photo-z* atribuído aos objetos no momento do ajuste de modelos não causa impactos positivos sobre esse resultado.

Tabela 5.6: Classificação das estrelas e galáxias dos catálogos DC6B\_S e DC6B\_G segundo o menor  $\chi^2$  obtido com o LePhare.

Classe	Amostra de estrelas	Amostra de galáxias
estrela	1288738 (26, 26%)	331894 (11, 17%)
galáxia	398555 (30, 92%)	1146912 (38, 60%)
quasar	551808 (42, 82%)	1492328 (50, 23%)

## 5.3 Photo-z

A aplicação dos dois códigos de *photo-z*, ANNz e LePhare, foi realizada com o catálogo DES\_66. Optamos por não utilizar o catálogo DES\_5000 pois o catálogo menor é representativo da simulação da área total do céu.

### 5.3.1 Redes Neurais

Para a aplicação do ANNz, separamos 10% do catálogo DES\_66 como amostra de treinamento e outros 10% como amostra de validação. As iterações de treinamento foram realizadas até que o código afirmasse ter atingido a convergência. Na figura 5.4 mostramos a comparação entre o *spec-z* e o *photo-z*. As aglomerações em torno de  $z_{photo} \approx 3,3$  e  $z_{photo} \approx 3,9$  e a partir de  $z_{photo} \approx 4,1$  mostram a dificuldade do ANNz em lidar com os *dropouts* gerados pelo AddQSO nas bandas mais azuis. Nessas faixas de *redshift* há a presença de *g-dropouts* e *r-dropouts* que assumem valores 99,999 que devem dificultar o treinamento e posterior cálculo de *photo-z* do



código. Um outro efeito que pode ser observado na figura 5.4 é a piora do ajuste para  $z_{espec} < 1,5$ , ou seja, mais perto da borda de *redshift* da amostra de treinamento.

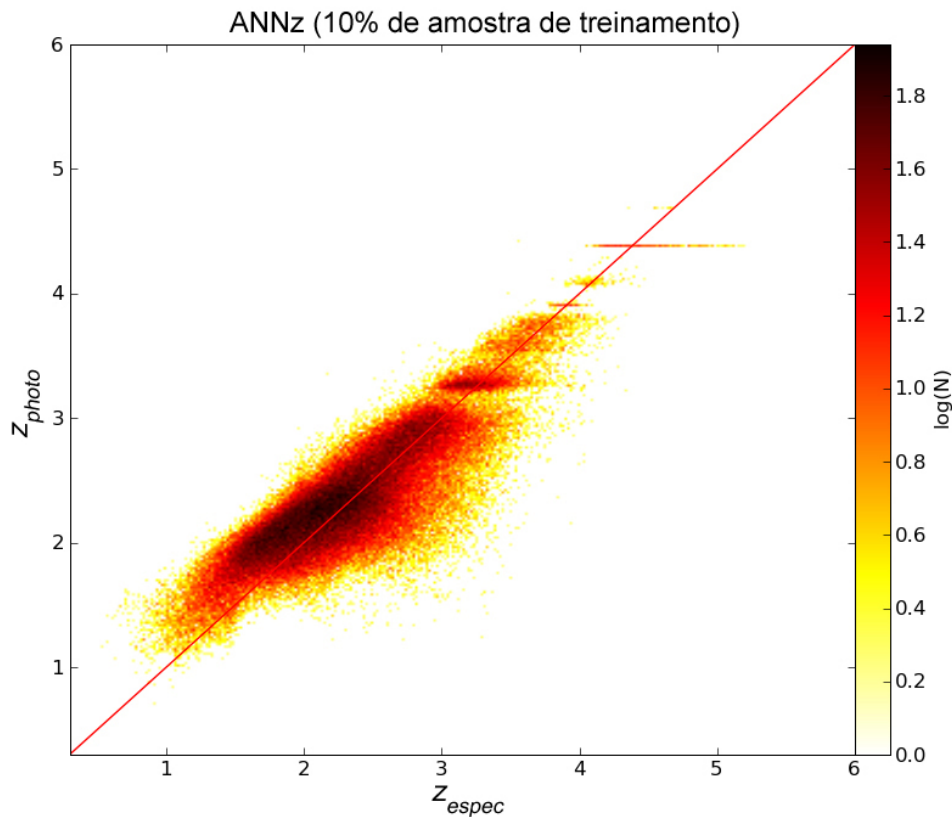


Figura 5.4: Comparação entre  $z_{photo}$  e  $z_{espec}$ . Resultado obtido com o código ANNz para o catálogo DES\_66.

As métricas para o cálculo de *photo-z* com o ANNz estão listadas na tabela 5.7. Os resultados mostram apenas 12,4% de identificações catastróficas. As outras métricas mostram um resultado comparável àquele obtido na seção 3.1 para os quasares do QDR7 na presença da banda *u*. Isso mostra que, para o algoritmo ANNz, a falta dessa banda não é tão crítica e é compensada pelo acréscimo da banda *Y*.

O ótimo resultado foi obtido com uma amostra de treinamento relativamente grande. Entretanto, para a área do DES é provável que se tenha, no máximo, 100000 quasares com espectroscopia para formarem uma amostra de treinamento. Se considerarmos que nossa previsão são de  $\approx 10^7$  quasares nos 5000 graus quadrados do DES, a melhor amostra de treinamento será composta por apenas 1% dos objetos. Sendo assim, refizemos o teste com o ANNz, diminuindo as amostra de treinamento e validação para apenas 1% do catálogo DES\_66. A comparação entre *photo-z* e

$espec-z$  pode ser vista na figura 5.5 e as métricas estão listadas na tabela 5.7.

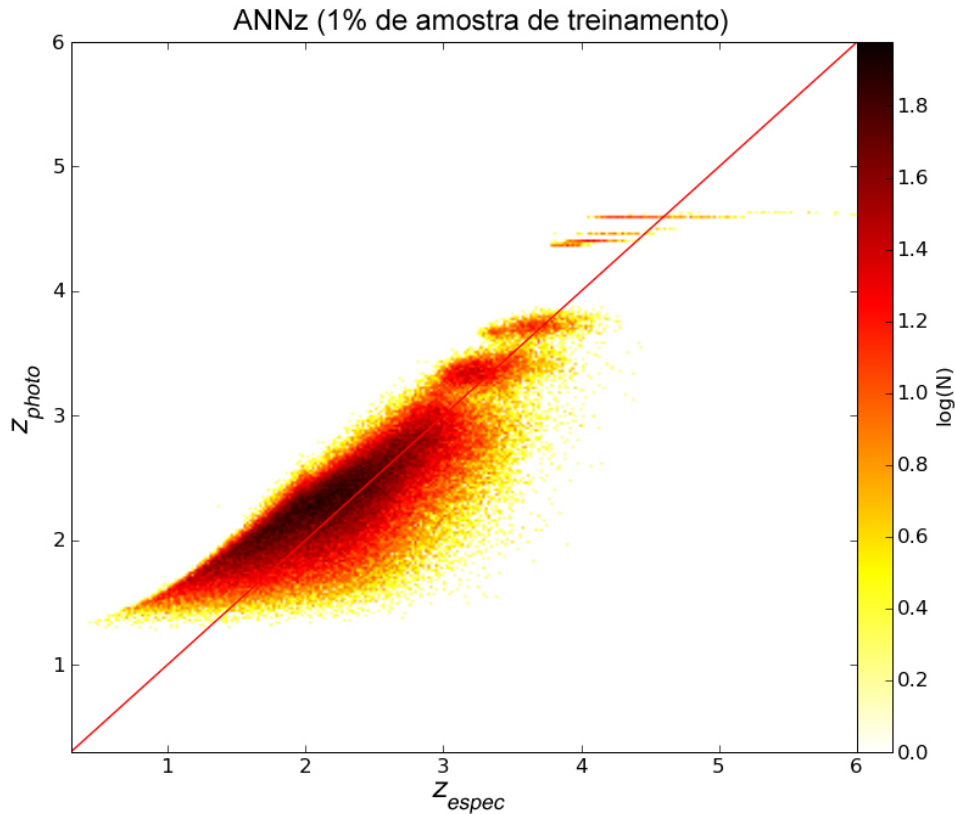


Figura 5.5: Comparação entre  $z_{photo}$  e  $z_{espec}$ . Resultado obtido com o código ANNz para o catálogo DES\_66 com amostras de treinamento e validação compostas por 1% da amostra alvo.

Tabela 5.7: Métricas do resultado do cálculo de photo-z do ANNz para o catálogo simulado pelo AddQSO com diferentes tamanhos de amostra de treinamento.

	10%	1%
bias	0,005	-0,005
$\sigma$	0,337	0,362
$\sigma_{68}$	0,300	0,329
Fração de $id_{catast.}$	0,124	0,158

Como podemos observar, há uma piora no resultados dos objetos que apresentam  $z_{espec} < 1,5$ , o que se traduz num aumento dos valores de  $\sigma$ ,  $\sigma_{68}$  e na fração de  $id_{catast.}$ . Apesar disso, o impacto no resultado não foi muito grande.

### 5.3.2 Ajuste de Modelos

A comparação entre  $espec-z$  e o  $photo-z$  obtido com o LePhare para o catálogo DES\_66 é exibida na figura 5.6. Pode-se observar que há acúmulos de pontos em

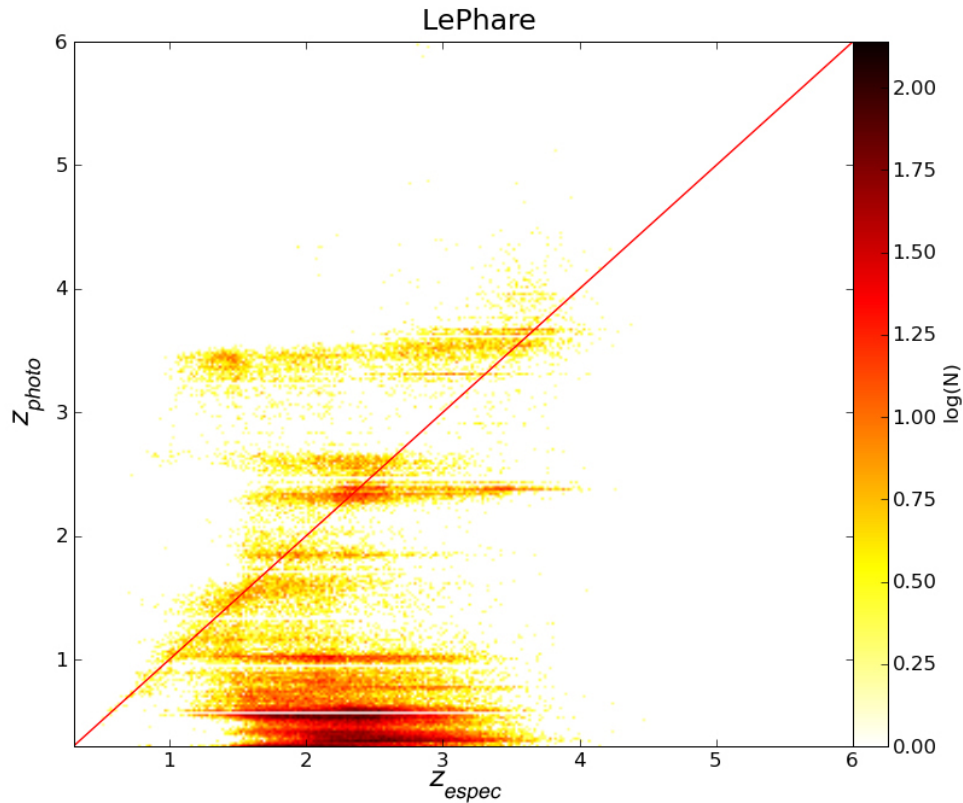


Figura 5.6: Comparação entre  $z_{photo}$  e  $z_{espec}$ . Resultado obtido com o código LePhare para o catálogo DES\_66.

linhas horizontais, que são fruto da discretização dos valores de *photo-z* calculados pelo LePhare, que são atribuídos em intervalos de *redshift* de 0,02. Há uma queda significativa na qualidade dos resultados principalmente na região onde  $z_{espec} < 3,2$ , que ainda não apresenta a linha de Ly $\alpha$  em nenhuma das bandas do DES. As falhas no cálculo de *photo-z* para esses objetos eram esperadas de acordo com o que foi observado no resultado com os quasares do DR7 na seção 3.2, principalmente na ausência da banda *u*. A degenerescência de cor, que faz com que objetos de cores similares apresentem *redshifts* diferentes, parece ganhar peso sem a informação dessa banda mais azul.

Tabela 5.8: Métricas do resultado do cálculo de *photo-z* do LePhare para o catálogo DES\_66.

bias	-1,538
$\sigma$	1,121
$\sigma_{68}$	2,131
Fração de $id_{catast.}$	0,878

As métricas do resultado podem ser vistas na tabela 5.8. O resultado obtido

apresenta  $\approx 88\%$  de identificações catastróficas. A fração de  $id_{catast.}$  diminui para 0,716 caso selecionemos apenas objetos que tem erros fotométricos inferiores a 0,05 mag, porém permanece insatisfatória.

# Capítulo 6

## Conclusões e Perspectivas

Grandes levantamentos como o DES irão aumentar a amostra de quasares em todas as faixas de *redshift* permitindo um amplo estudo da demografia desses objetos e das condições físicas do meio intergaláctico. As ferramentas de seleção e análise devem ser testadas e preparadas para lidar com uma grande quantidade de dados. Com o objetivo de fazer previsões para o levantamento DES e simular catálogos capazes de testar essas ferramentas, desenvolvemos a suíte de algoritmos AddQSO.

Para validarmos os algoritmos do AddQSO, simulamos quasares em uma área de 8417 graus quadrados com magnitudes nos filtros SDSS. Comparamos o catálogo simulado com o catálogo de quasares espectroscopicamente confirmados do DR7 do SDSS, o QDR7, e com o catálogo de  $\approx 1,1 \times 10^6$  candidatos a quasares do SDSS, o QR09. A partir de nossos testes, podemos concluir:

- a distribuição de cores simuladas pelo AddQSO reproduz a distribuição de cores observadas no QDR7;
- há uma correlação entre a distribuição de cores com o *redshift*, principalmente para valores menores de  $z$ , entre os quasares simulados e os do QDR7. Em regimes de *redshifts* maiores há uma discrepância nos diagramas cor-cor, o que, pelo menos em parte, pode ser explicado pela baixa amostragem de objetos com  $z > 3,0$  no catálogo QDR7;
- o catálogo simulado pelo AddQSO subestima a quantidade de quasares no regime de menor magnitude aparente quando comparado ao catálogo QR09. Isso pode ser causado em parte pela contaminação por objetos como galáxias

Seyfert e estrelas, e também pela diferença na calibração das magnitudes do catálogo de QR09.

A simulação de um catálogo com 5000 graus quadrados de área e magnitudes nos filtros DES nos permitiu fazer previsões para o levantamento. As principais conclusões são:

- o DES será capaz de detectar em pelo menos uma de suas cinco bandas mais de  $10^7$  quasares;
- o quasar mais distante a ser observado terá seu *redshift* da ordem de 7,5;
- apesar da baixa profundidade da banda Y, o DES será capaz de observar em torno de 40 objetos com *redshift* superior a 7,0 e quase 1000 objetos com  $z > 6,0$ .

Apesar dos números esperados pelo DES, é necessário sermos capazes de identificar e classificar esses objetos corretamente. Para os testes dos métodos de classificação, utilizamos catálogos simulados para quasares, estrelas e galáxias em áreas similares. Os catálogos de estrelas e galáxias foram simulados pela colaboração do DES. Testamos a possibilidade de separar as três amostras utilizando diagramas cor-cor e a possibilidade de aplicar um método de ajuste de modelos, o LePhare, como classificador. As conclusões obtidas foram:

- apenas uma pequena fração dos quasares pode ser separada através da aplicação de limites de cor às amostras;
- com o método de classificação por ajuste de modelos, a completeza de um catálogo de candidatos a quasares é superior a 0,92. A taxa de contaminação, por outro lado, é muito alta, resultando em uma pureza de apenas 0,06 quando levamos em consideração a contaminação por estrelas e galáxias, e de 0,2 quando consideramos apenas estrelas como contaminantes.

Candidatos a quasares com grandes *redshifts* terão *follow-up* espectroscópico. No entanto, se nossas previsões de densidade de objetos para o DES se confirmarem, não será possível realizar espectroscopia de todos os quasares observados. Sendo assim, a maioria dos candidatos selecionados, principalmente aqueles mais próximos,

terão apenas valores de *photo-z*. Nesse trabalho, testamos a eficiência de dois códigos: ANNz, baseado em redes neurais, e LePhare, de ajuste de modelos. Baseado num conjunto de métricas que comparam o *photo-z* obtido com o *spec-z* proveniente da tabela verdade da simulação do AddQSO, podemos concluir:

- o código ANNz apresenta um ótimo desempenho. As identificações catastróficas são de, no máximo, 12,4%;
- o ANNz apresenta, para os quasares simulados, um desempenho comparável ao obtido com os quasares do QDR7, o que mostra que a ausência da banda *u* não é de grande impacto no resultado de *photo-z* na presença da banda *Y*;
- o código LePhare apresenta um desempenho insatisfatório, com as identificações catastróficas podendo chegar a 88% quando não são aplicados limites aos erros fotométricos;
- os testes do LePhare com os quasares do QDR7 mostram identificações catastróficas em 53,1% dos casos quando a banda *u* é removida, o que evidencia a fragilidade do método nessas condições.

Nossas perspectivas se concentram no AddQSO e nas ferramentas de classificação de objetos. No primeiro caso, planejamos aplicar melhorias visando simulações de quasares mais versáteis e realistas. Para atingir esse objetivo, iremos:

- incluir os filtros do infravermelho do VISTA de forma a termos uma melhor estatística da quantidade de objetos de grandes *redshifts* possíveis de serem observados e podermos aprimorar as ferramentas para classificá-los.
- disponibilizar conjuntos de filtros de outros levantamentos;
- incluir outras QLFs, como a de Croom et al. (2009) e a de Palanque-DeLabrouille et al. (2012) no regime de pequenos *redshifts*, a de Fontanot et al. (2007) para grandes *redshifts* e a QLF com dados do DR9 do SDSS que deve ser publicada em breve;
- estudar a possibilidade de acrescentar uma QLF de AGNs de tipo 1, como a de Bongiorno et al. (2007);

- desenvolver um método de simulação de erros fotométricos gaussianos, para podermos aplicar erros a filtros que não sejam do DES e que, portanto, não temos uma distribuição na qual nos basearmos.
- aplicar a distribuição de erros de fontes pontuais do SDSS para acrescentarmos erros às simulações realizadas com esse conjunto de filtros;
- buscar um modelo de halos de quasares que apresente uma forma de posicionar esses objetos de acordo com a distribuição de matéria escura, de forma similar ao que é feito nas simulações de galáxias da colaboração DES para termos maior coesão e possamos criar um catálogo de fontes observáveis no DES que seja coerente com a distribuição de matéria.

Na questão da classificação de objetos, nosso foco será a implementação e o teste de outras ferramentas de classificação como:

- XDQSO (Bovy et al. 2010), uma ferramenta desenvolvida para selecionar quasares do SDSS com *redshifts* intermediários ( $z \approx 2, 5$ );
- algoritmos de agrupamento como o agrupamento por k-Médias (e.g., MacQueen 1967);
- redes neurais.

O XDQSO também pode ser aplicado no cálculo de *photo-z*. Já no caso das duas últimas técnicas, podemos usar, além das cores, outros parâmetros como índice de estelaridade e até mesmo uma classificação prévia obtida por outro método como parâmetros para definir cada grupo. Para escolher os parâmetros relevantes a serem utilizados nas definições dos grupos, podemos aplicar uma análise de componentes principais.

Por último, iremos aplicar as ferramentas de classificação e cálculo de *photo-z* testadas nesse trabalho aos dados reais obtidos pelo DES. Os dados que resultaram do período de verificação científica estarão disponíveis em breve.



# Apêndice A

## Códigos XML do AddQSO

### A.1 Estrutura do *workflow* do AddQSO

---

```
<?xml version="1.0" encoding="windows-1250"?>
<pipeline name="AddQSO">
  <taskgroup id="addqso" desc="Generation of a QSO sample">
    <task id="addqsolf" name="Integration of QLF" config="yes" confName="addqsolf">
      <components>
        <component id="addqsolf" />
      </components>
    </task>
    <task id="addqsomag" name="Magnitude Generation" config="yes">
      <components>
        <component id="addqsomag" />
      </components>
      <dependencies>
        <dependency id="addqsolf">
          <file mimetype="text/table" class="science"/>
        </dependency>
      </dependencies>
    </task>
    <task id="addqsoplotmag" name="Plots" config="no">
      <components>
        <component id="addqsoplotmag" />
      </components>
      <dependencies>
        <dependency id="addqsomag">
          <file mimetype="text/table" class="science"/>
        </dependency>
      </dependencies>
    </task>
  </taskgroup>
</pipeline>
```

```

    <config type="string" id="filter" />
  </dependency>
</dependencies>
</task>
<task id="addqsoerrors" name="Photometric Errors" config="no">
  <components>
    <component id="addqsoerrors" />
  </components>
  <dependencies>
    <dependency id="addqsomag">
      <file mimetype="text/table" class="science"/>
      <config type="string" id="filter" />
    </dependency>
  </dependencies>
</task>
</taskgroup>
</pipeline>

```

---

## A.2 Configuração do módulo de integração da QLF

---

```

<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="/static/xml/mkform.xsl"?>

<config module="addqsolf" version="001">
  <section id="Luminosity Function">
    <subsection>
      <scalar id="LF" name="Luminosity Function" type="int" default="1">
        <values>
          <value value="1">Richards et al. (2006)</value>
          <value value="2">Willott et al. (2010)</value>
          <value value="3">Richards + Willott</value>
        </values>
      </scalar>
    </subsection>
  </section>
  <section id="Cosmological Parameters">
    <subsection>

```

```
<scalar default="0.3" id="OMGM" name="Density parameter (matter)"
  type="float">
  <values>
    <value>0.3</value>
  </values>
</scalar>
<scalar default="0.7" id="OMGL" name="Density parameter (lambda)"
  type="float">
  <values>
    <value>0.7</value>
  </values>
</scalar>
<scalar default="-1.0" id="W" name="w" type="float">
  <values>
    <value>-1.0</value>
  </values>
</scalar>
<scalar default="0.7" id="HPAR" name="Hubble parameter" type="float">
  <values>
    <value>0.7</value>
  </values>
</scalar>
</subsection>
</section>
<section id="QLF Integration Parameters">
  <subsection>
    <scalar default="0.3" id="ZMIN" name="Minimum redshift" type="float">
      <values>
        <value>0.0</value>
      </values>
    </scalar>
    <scalar default="5.3" id="JOINZ" name="QLF change redshift (if Richards +
      Willott is selected)" type="float">
      <values>
        <value>5.3</value>
      </values>
    </scalar>
    <scalar default="8.0" id="ZMAX" name="Maximum redshift" type="float">
      <values>
```

```
        <value>8.0</value>
      </values>
    </scalar>
    <scalar default="0.05" id="DZ" name="Redshift bin" type="float">
      <values>
        <value>0.05</value>
      </values>
    </scalar>
    <scalar default="-30.0" id="MAGMIN" name="Minimum absolute i magnitude
      (bright)" type="float">
      <values>
        <value>-30.0</value>
      </values>
    </scalar>
    <scalar default="-23.5" id="MAGMAX" name="Maximum absolute i magnitude
      (faint)" type="float">
      <values>
        <value>-23.5</value>
      </values>
    </scalar>
    <scalar default="0.05" id="DM" name="Absolute magnitude bin" type="float">
      <values>
        <value>0.05</value>
      </values>
    </scalar>
    <scalar default="5000.0" id="ADEG" name="Simulated area (square degree)"
      type="float">
      <values>
        <value>5000.0</value>
      </values>
    </scalar>
  </subsection>
</section>
</config>
```

---

# Apêndice B

## Atribuição de erros

### B.1 A adição de $\Delta m$ às bandas $g$ , $r$ , $z$ , e $Y$

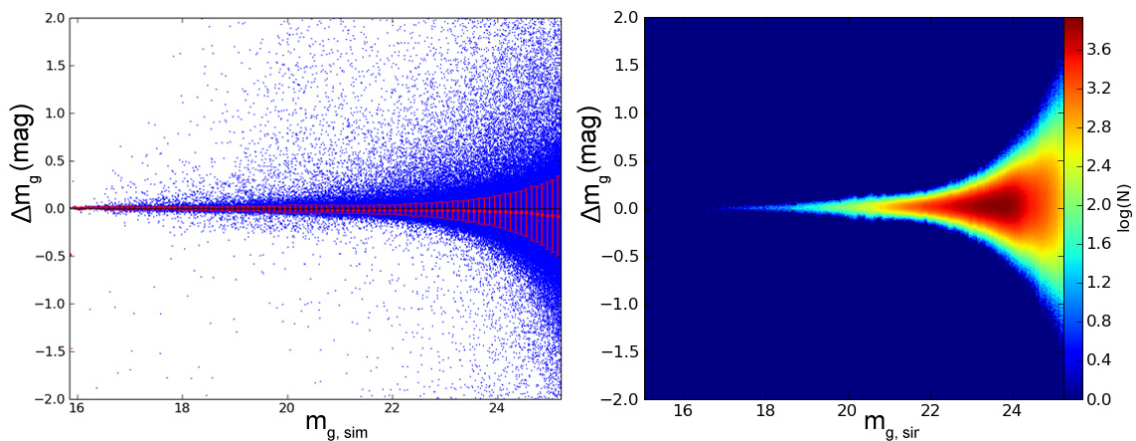


Figura B.1: Distribuição de  $\Delta m_g$  em função da magnitude simulada na banda  $g$ . No painel da esquerda vemos em azul a distribuição de  $\Delta m_g$  para as estrelas do DC6B e, sobrepostas em vermelho, as médias com barras de  $1\sigma$  calculadas após a remoção dos objetos que apresentavam  $\Delta m_g$  fora do limite de  $3\sigma$ . No painel da direita está a distribuição de  $\Delta m_g$  simulada para quasares em uma área de 5000 graus quadrados no filtro  $g$  do DES. A barra colorida representa o logartimo da contagem de objetos em cada ponto.

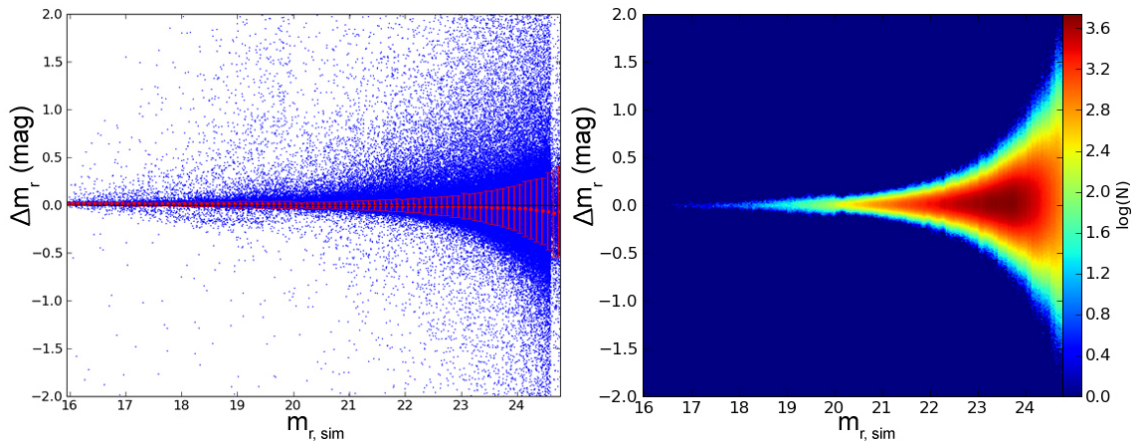


Figura B.2: O mesmo que na figura B.1 para a banda  $r$ .

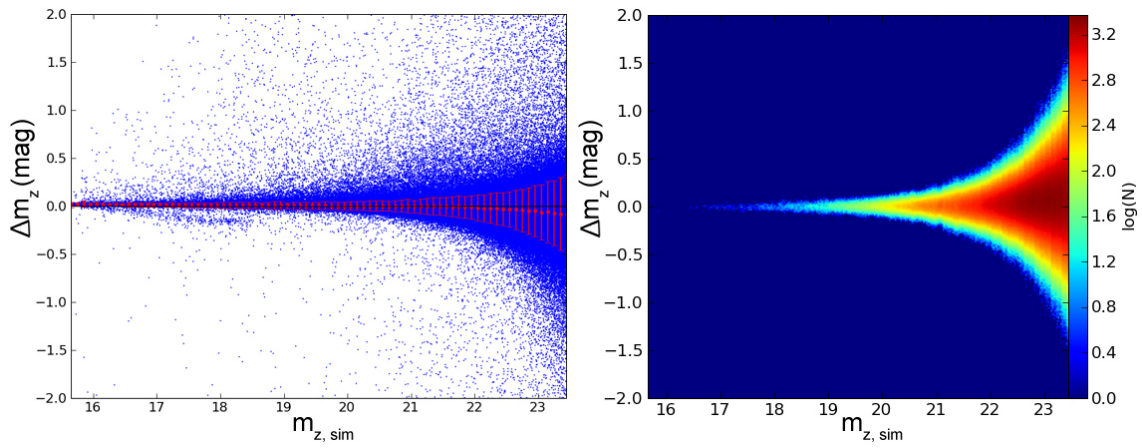


Figura B.3: O mesmo que na figura B.1 para a banda  $z$ .

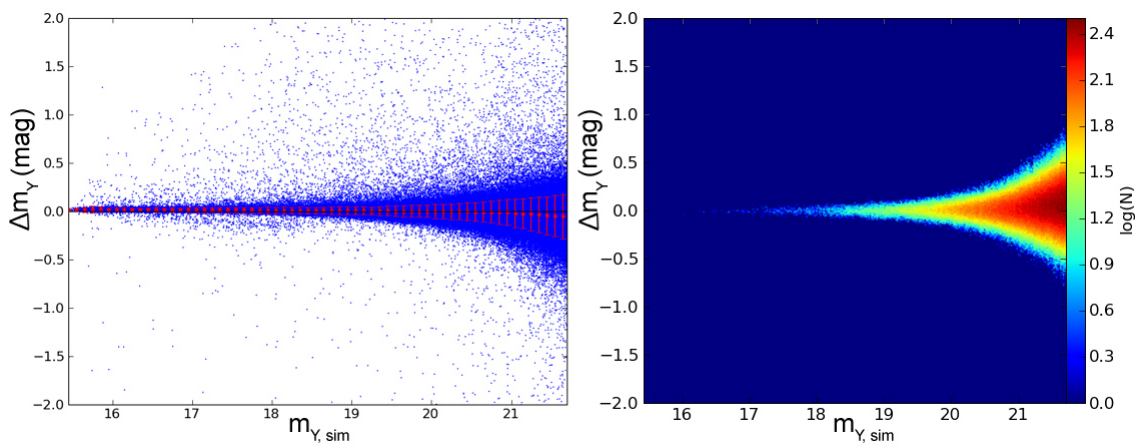


Figura B.4: O mesmo que na figura B.1 para a banda  $Y$ .

## B.2 A adição de incertezas às bandas $g$ , $r$ , $z$ , e $Y$

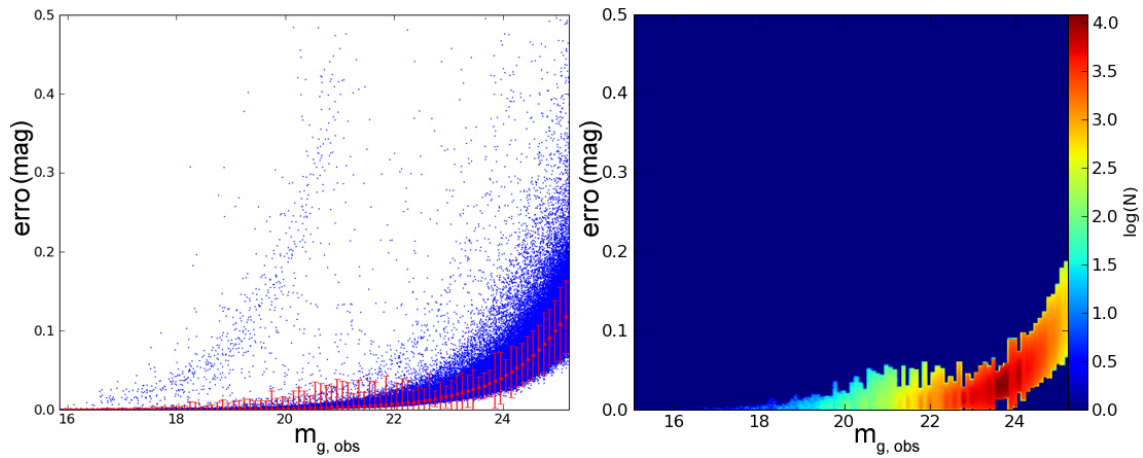


Figura B.5: Erros fotométricos na banda  $g$  em função da magnitude observada na mesma banda. No painel da esquerda vemos em azul a distribuição de erro para as estrelas do DC6B e, sobrepostas em vermelho, as medianas e as barras de  $1\sigma$ . No painel da direita está a distribuição de erros fotométricos simulada para quasares em uma área de 5000 graus quadrados no filtro  $g$  do DES. A barra colorida representa o logaritmo da contagem de objetos em cada ponto.

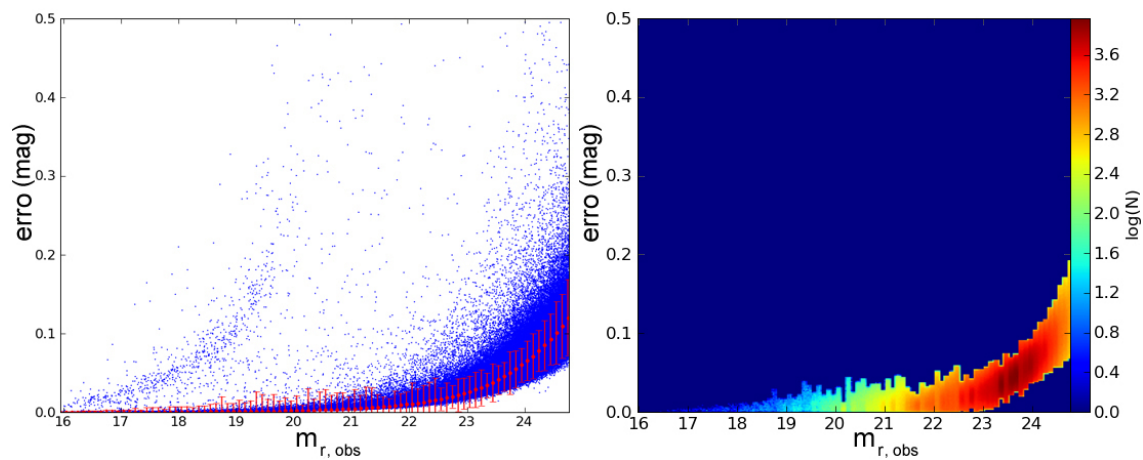


Figura B.6: O mesmo que na figura B.5 para a banda  $r$ .

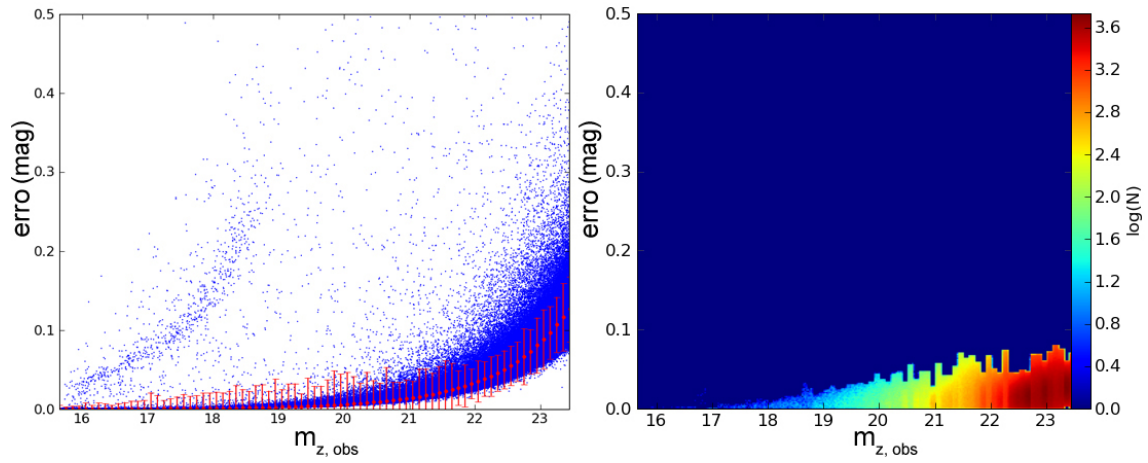


Figura B.7: O mesmo que na figura B.5 para a banda  $z$ .

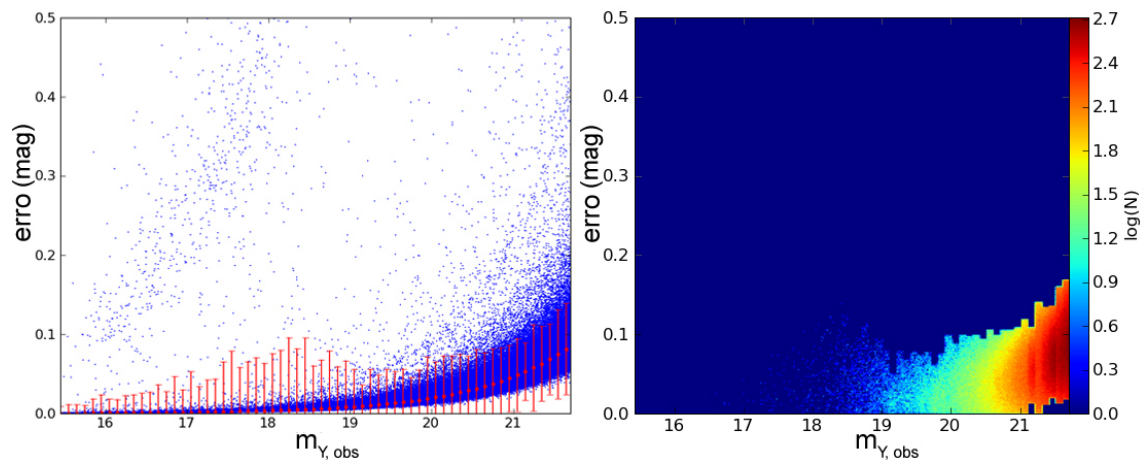


Figura B.8: O mesmo que na figura B.5 para a banda  $Y$ .



# Referências

- Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A. et al. 2005, AJ 129, 1755
- Abramo, L. R., Strauss, M. A., Lima, M. et al. 2012, MNRAS 423, 3251
- Antonucci, R. 1993, ARAA 31, 473
- Abbott, T., Aldering, G., Annis, J. et al. 2005, arXiv:astro-ph/0510346
- Annis, J., Bridle, S., Castander, F. J. et al. 2005, arXiv:astro-ph/0510195v1
- Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS 310, 540
- Becker, G. D., Rauch, M., Sargent, W. L. W. 2007, ApJ 662, 72
- Bennert, N., Falcke, H., Schulz, H. et al. 2002, ApJ 574, L105
- Bennett, A. S. 1962, *MemRAS* 68, 163
- Bertin, E. & Arnouts, S. 1996, AAPS 317, 393
- Bohlin, R. C., Colina, L. & Finlay, D. S. 1995, AJ 110, 1316
- Bongiorno, A., Zamorani, G., Gavignaud, I. et al. 2007, A&A 472, 443
- Boutsia, K., Grazian, A., Giallongo, E. et al. 2011, ApJ 736, 41
- Bovy, J., Hennawi, J. F., Hogget, D. W. et al. 2010, ApJ 729, 141
- Boyle, B. J., Shanks, T., Croom, S. M. et al. 2000, MNRAS 317, 1014
- Budavári, T., Csabai, I., Szalay, A. S., et al. 2001, AJ 122, 1163
- Bundy, K., Georgakakis, A., Nandra, K. et al. 2008, ApJ 681, 931
- Calzetti, D., Kinney, A. L., Storchi-Bergmann, T. 1994, ApJ 429, 582
- Canalizo, G., Bennert, N., Jungwiert, B. et al. 2007, ApJ 669, 801
- Capetti, A., Axon, D. J., Macchetto, F. et al. 1996, ApJ 469, 554
- Coleman, G. D., Wu, C.-C., Weedman, D. W. 1980, ApJS 43, 393
- Collister, A. & Lahav, O. 2004, PASP 116, 818
- Croom, S. M., Richards, G. T., Shanks, T. et al. 2009, MNRAS 399, 1755
- Croom S. M., Smith R. J., Boyle B. J., et al. 2001, MNRAS 322, L29
- Croom S. M., Smith R. J., Boyle B. J., et al. 2004, MNRAS 349, 1397

- Dunkley, J., Spergel, D. N., Komatsu, E. et al. 2009, ApJ 701, 1804
- Edge, D. O., Shakeshaft, J. R., McAdam, W. B., Baldwin, J. E., Archer, S. 1959, *MemRAS* 68, 37
- Fan, X., Strauss, M. A., Schneider, D. P., et al. 2001, AJ 121, 54
- Fan, X., Narayanan, V. K., Lupton, R. H. et al. 2001, AJ 122, 2833
- Fan, X., Strauss, M. A., Richards, G. et al. 2006, AJ 131, 1203
- Fan, X., Strauss, M. A., Schneider, D. P. et al. 2003, AJ 125, 1649
- Firth, A., Lahav, O., Somerville, R. 2003, MNRAS 339, 1195
- Fontanot, F., Cristiani, S., Monaco, P. et al. 2007, A&A 461, 39
- Fukugita, M., Hogan, C. J., Peebles, P. J. E. 1998, ApJ 503, 518
- Giacconi, R., Zirm, A., Wang, Ju. et al. 2002, ApJS 139, 369
- Greenstein, J. L., Matthews, T. A. 1963, AJ 68, 280
- Gregg, M. D., Lacy, M., White, R. L. et al. 2002, ApJ 564, 133
- Gunn, J. E. & Peterson, B. A. 1965, AJ 142, 1633
- Hatziminaoglou, E., Mathez, G., Pelló, R. 2000, A&A 359, 9
- Hatziminaoglou, E., Pérez-Fournon, I., Polletta, M. et al. 2005, AJ 129, 1198
- Hong-Yee, C. 1964, Phys. Today 17
- Hornschemeier, A. E., Brandt, W. N., Garmire, G. P. et al. 2000, ApJ 541, 49
- Ilbert, O., Arnouts, S., McCracken, H.J., et al. 2006, A&A 457, 841
- Komatsu, E., Smith, K. M., Dunkley, J. et al. 2010, ApJS 192, 18
- Kormendy, J. & Richstone, D. 1995, ARA&A 33, 581
- Lamastra, A., Menci, N., Maiolino, R. et al. 2010, MNRAS 405, 29
- Lonsdale, C. J., Smith, H. E., Rowan-Robinson, M. et al. 2003, PASP 115, 897
- Lynden-Bell, D. 1969, Nature 223, 690
- MacQueen, J. B. *Some Methods for classification and Analysis of Multivariate Observations*. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, 1967. p. 281 – 297.
- Magorrian, J., Tremaine, S. Richstone, D. et al. 1998, AJ 115, 2285
- Martini, P. (2004), *Carnegie Observatories Astrophysics Series, Vol. 1: Coevolution of Black Holes and Galaxies*. New York: Cambridge University Press.
- Matthews, T. A., Sandage, A. R. 1963, AJ 138, 30
- McMahon, R. G. & Irwin, M. 1991, ASPCS 21, 391

- Mitra, S., Choudhury, T. R., Ferrara, A. 2011, MNRAS 413, 1569
- Mo, Houjun; van den Bosch, Frank; White, Simon (2010), *Galaxy Formation*. New York: Cambridge University Press.
- Mortlock, D. J., Patel, M., Warren, S. J. et al. 2008, arXiv:0810.3859v1
- Ono, Y., Ouchi, M., Mobasher, B. et al. 2012, ApJ 744, 83
- Osmer, P. S. 1982, ApJ 253, 28
- Palanque-Delabrouille, N., Magneville, Ch., Yèche, Ch. et al. 2012, A&A 551, 29
- Polletta, M., Wilkes, B. J., Siana, B. et al. 2006, ApJ 642, 673
- Polletta, M., Tajer, M., Maraschi, L. et al. 2007, ApJ 663, 81
- Raban, D., Jaffe, W., Röttgering, H. et al. 2009, MNRAS 394, 1325
- Richards, G. Fan, X., Newberg, H. J. et al. 2002, AJ 123, 2945
- Richards, G. T., Myers, A. D., Gray, A. G. et al. 2009, ApJS 180, 67
- Richards, G. T., Strauss, M. A., Fan, X. et al. 2006, AJ 131, 2766
- Ruhl, J. E., Ade, P. A. R., Carlstrom, J. E. et al. 2004, SPIE 5498, 11
- Salpeter, E. E. 1964, ApJ 140, 796
- Salvato, M., Hasinger, G., Ilbert, O. et al. 2009, ApJ 690, 1250
- Schmidt, M. 1963, Nature 197, 1040
- Schmidt, M. & Green, R. F. 1983, ApJ 269, 352
- Schmidt, M., Schneider, D. P., Gunn, J. E. 1995, AJ 110, 68
- Schneider, D. P., Hall, P. B., Richards, G. T. et al. 2007, AJ 134, 102
- Schneider, D. P., Richards, G. T., Hall, P. B. et al. 2010, AJ 139, 2360
- Schneider, Peter (2006), *Extragalactic astronomy and cosmology : an introduction*. Berlin: Springer.
- Shankar, F. & Mathur, S. 2007, ApJ 660, 1051
- Songaila, A. 2004, AJ 127, 2598
- Sparke, Linda & Gallagher, John (2007), *Galaxies in the Universe: an introduction*. New York: Cambridge University Press.
- Pickles, A. 1998, PASP 110, 863
- Tristram, K. R. W., Meisenheimer, K., Jaffe, W. et al. 2007, A&A 474, 837
- Ueda, Y., Akiyama, M., Ohta, K. & Miyaji, T. 2003, ApJ 598, 886
- Venemans, B. P., McMahon, R. G., Warren, S. J. et al. 2007, MNRAS 376, L76
- Volonteri, M. & Rees, M.J. 2005, ApJ 633, 624

Warren, S. J., Hewett P. C., Osmer P. S. 1994, *ApJ* 421, 412

Weisskopf, M. C., Tananbaum, H. D., van Speybroeck, L. P., & O'dell, S. L. 2000, *Proc. SPIE* 4012, 2

Willott, C. J., Delorme, P., Omont, A. et al. 2007, *AJ* 134 2435

Willott, C. J., Delorme, P., Reylé, C. et al. 2009, *AJ* 137, 3541

Willott, C. J., Delorme, P., Reylé, C. et al. 2010, *AJ* 139, 918

York, D. G., Adelman, J., Anderson, J. E. Jr. et al. 2000, *AJ* 120, 1579

Yèche, Ch., Petitjean, P., Rich, J. et al. 2010, *A&A* 523, A14

Zel'dovich, Y. B., Novikov, I. D. 1964, *Sov. Phys. Dokl.* 158, 811