

**Universidade Federal do Rio de Janeiro**  
**Centro de Ciências Matemáticas e da Natureza**  
**Observatório do Valongo**  
**Departamento de Astronomia**

Estudo preliminar da qualidade de *redshifts* fotométricos  
para o Dark Energy Survey

Luiz Felipe Melo de Menezes

Orientador: Dr. Paulo Sérgio de Souza Pellegrini (Observatório Nacional)

**Projeto de Conclusão de Curso para a obtenção do título de Astrônomo**

Rio de Janeiro  
2016

## **Agradecimentos**

Inicialmente gostaria de agradecer a todos que, de alguma forma, ajudaram a tornar este momento possível.

À minha família, em especial aos meus pais, que sempre me incentivaram e me deram tranquilidade para fazer o curso. E claro, também ao meu irmão, que com sua inocência e seu carinho, consegue me fazer sorrir até nos momentos de tensão.

À minha namorada, que admira tanto a minha profissão e fez com que os últimos meses fossem os melhores da minha vida. Juliana, amor, você é uma pessoa incrível e uma profissional exemplar, obrigado por todo o apoio.

Aos meus eternos amigos dos tempos de colégio. Felipe, Lucas, Mariana, Natália e Wesley, obrigado por serem tão especiais, cada um à sua maneira. Quero que estejam sempre presentes em minha vida.

Aos meus amigos da Astronomia, que estiveram comigo nessa caminhada e com os quais tive vários bons momentos. Desejo que todos tenham uma carreira de muito sucesso.

Aos amigos que fiz em outros cursos, sobretudo Bruno, Catherine, Claudio e Mariana. Juntos tivemos momentos de aflição estudando para as provas, mas muito mais frequentes foram os momentos de risadas e descontração. No fim das contas, nos divertimos bastante e deu tudo certo.

Aos professores que tive ao longo da vida, tanto os da época do colégio quanto os da UFRJ. No OV, no IF e no IM tive contato com profissionais do mais alto nível, os quais demonstravam uma paixão contagiante durante as aulas e faziam de tudo para que nós alunos aprendêssemos as disciplinas da melhor forma possível.

À equipe de funcionários do OV, em especial a Rosa e a Claudia. Sempre empenhadas em resolver os problemas administrativos, me ajudaram muito todas as vezes que precisei.

Ao ON e ao coordenador do LIneA, que me deram a oportunidade de trabalhar nesse projeto, e ao CNPq, que tornou tudo possível ao me auxiliar com uma bolsa de iniciação científica.

Por fim, àquela que foi a pessoa mais importante na minha jornada rumo ao diploma de Astrônomo, meu orientador na iniciação científica, Paulo Pellegrini. Sempre compreensivo e atencioso, você me ajudou não somente no projeto, como também com conversas e conselhos sobre a área. Agradeço por tudo, e te desejo o melhor sempre.

## Resumo

Esta monografia contém uma primeira avaliação da qualidade dos *redshifts* fotométricos (photo-zs) obtidos com dois códigos diferentes, sendo um deles de ajuste de *templates* (LePhare) e o outro de treinamento (TPZ), no contexto do que se espera para os dados do Dark Energy Survey (DES). Em particular, o LePhare foi mais explorado em função de sua capacidade de fornecer parâmetros físicos das galáxias.

Resultados são mostrados para a avaliação do aumento dos erros tanto nos *redshifts* fotométricos quanto nos limites inferiores de incertezas para magnitudes absolutas determinadas, devido à ausência da banda u nos dados do DES.

O código LePhare produz resultados com incertezas para o photo-z de  $\sigma_{68} = 0.08$ , sendo mais confiáveis acima de photo-zs = 0.6. O limite inferior para incertezas na magnitude absoluta é de 0.45 mag.

Por outro lado, verifica-se que o TPZ produz photo-zs de melhor qualidade. Utilizando-se esse código, em dados com sistema de 5 filtros de banda larga grizY, a incerteza nas determinações de photo-zs é de cerca de  $\sigma_{68} = 0.05$  e um limite inferior para a incerteza na magnitude absoluta é 0.25 mag.

Uma alternativa de utilização combinada dos dois códigos é mostrada como uma boa opção de determinação de parâmetros relevantes para o estudo de propriedades de galáxias.

**Palavras-Chave:** *redshifts* fotométricos, magnitudes absolutas, galáxias, incertezas, códigos de photo-z, LePhare, TPZ

## Abstract

This dissertation contains an initial assessment of the quality of photometric redshifts (photo-zs) obtained by two different codes: a template setting method (LePhare) and a training method (TPZ). In particular, LePhare was further explored due to its ability in providing physical parameters of galaxies.

Results are shown for the evaluation of the increase of errors in photometric redshifts and in the lower limit of uncertainty for absolute magnitudes due to the absence of the u band in DES data.

Code LePhare produces results with uncertainties of  $\sigma_{68} = 0.08$ , which are more reliable above photo-zs = 0.6. The lower limit for uncertainties in the absolute magnitude is 0.45 mag.

On the other hand, TPZ produces photo-zs of higher quality. Using this code on data with a grizY broadband 5 filters system, the uncertainty in the photo-zs determinations is about  $\sigma_{68} = 0.05$  and a lower limit for the uncertainty in the absolute magnitude is 0.25 mag.

A combination of both codes is shown to provide a good option for determining relevant parameters to the study of galaxy properties.

**Keywords:** photometric redshifts, absolute magnitudes, galaxies, uncertainties, photo-z codes, LePhare, TPZ

## Sumário

<b>1</b>	<b>Introdução .....</b>	<b>1</b>
<b>2</b>	<b>Métodos de determinação de photo-zs .....</b>	<b>5</b>
2.1	Classificação dos métodos .....	5
2.2	Descrição sucinta dos métodos .....	7
2.2.1	Dark Energy Survey Data Management (DESDM) .....	8
2.2.2	Artificial Neural Network (ANNz) .....	9
2.2.3	Trees for Photo-z (TPZ) .....	9
2.2.4	Relevance Vector Machine (RVMz).....	10
2.2.5	Normalized Inner Product – Nearest Neighbor (NIP-kNNz).....	11
2.2.6	Artificial Neural Network 2 (ANNz2) .....	12
2.2.7	ArborZ .....	12
2.2.8	Skynet .....	13
2.2.9	Bayesian Photometric Redshift (BPZ) .....	14
2.2.10	Photometric Analysis for Redshift Estimations (LePhare) .....	15
2.2.11	Zurich Extra-galactic Bayesian Redshift Analyser (ZEBRA) ..	17
2.2.12	PhotoZ.....	18
<b>3</b>	<b>Avaliação da performance de códigos de photo-zs .....</b>	<b>20</b>
3.1	LePhare .....	23
3.1.1	Photo-zs determinados com o LePhare .....	29
3.1.1.1	Padrão de referência .....	29
3.1.1.2	Amostra simulada do DES (DC6B) .....	32
3.1.1.3	Amostra VVDS sem a banda u .....	38
3.1.1.4	Amostra de Verificação Científica do DES .....	40
3.1.2	Magnitudes Absolutas determinadas com o LePhare .....	44
3.2	TPZ .....	47
3.2.1	Photo-zs determinados com o TPZ .....	50
3.2.1.1	Amostra VVDS sem a banda u .....	50
3.2.1.2	Amostra da Verificação Científica do DES .....	51
3.2.2	Magnitudes Absolutas no LePhare com photo-zs do TPZ.....	54
<b>4</b>	<b>Conclusões e comentários finais .....</b>	<b>56</b>
<b>5</b>	<b>Bibliografia .....</b>	<b>60</b>

## Lista de Figuras

<b>Figura 1:</b> SEDs CWW .....	24
<b>Figura 2:</b> SEDs Kinney .....	25
<b>Figura 3:</b> SEDs CWW deslocadas em $redshift = 0.2$ . .....	26
<b>Figura 4:</b> SEDs CWW deslocadas em $redshift = 1.0$ .....	26
<b>Figura 5:</b> SEDs de galáxias Elípticas, do CFHTLS, obtidas por interpolação.....	27
<b>Figura 6:</b> Comparação de zphotos obtidos no LePhare com zspecs do VVDS, usando SEDs CFHTLS. ....	30
<b>Figura 7:</b> Comparação de zphotos obtidos no LePhare com zspecs do VVDS, usando SEDs CFHTLS e com distinção dos tipos de galáxias .....	32
<b>Figura 8:</b> Comparação de zphotos obtidos no LePhare com zspecs do DC6B, usando SEDs CWW. ....	33
<b>Figura 9:</b> Comparação de zphotos obtidos no LePhare com zspecs do DC6B, usando SEDs CFHTLS.....	34
<b>Figura 10:</b> Comparação de zphotos obtidos no LePhare com zspecs do DC6B, usando SEDs CWW e com distinção dos tipos de galáxias.....	35
<b>Figura 11:</b> Comparação de zphotos obtidos no LePhare com zspecs do DC6B, usando SEDs CFHTLS e com distinção dos tipos de galáxias .....	36
<b>Figura 12:</b> Histograma dos zphotos obtidos no LePhare e zspecs do DC6B, usando SEDs CWW .....	37
<b>Figura 13:</b> Histograma dos zphotos obtidos no LePhare e zspecs do DC6B, usando SEDs CFHTLS.....	37

<b>Figura 14:</b> Comparação de zphotos obtidos no LePhare com zspecs do VVDS sem a banda u, usando SEDs CFHTLS .....	39
<b>Figura 15:</b> Comparação de zphotos obtidos no LePhare com zspecs do SVdeep, usando SEDs CFHTLS .....	41
<b>Figura 16:</b> Comparação de zphotos obtidos no LePhare com zspecs do SVdeep, usando SEDs CFHTLS e com distinção dos tipos de galáxias....	42
<b>Figura 17:</b> Histograma dos zphotos obtidos no LePhare e zspecs do SVdeep, usando SEDs CFHTLS e com distinção dos tipos de galáxias....	43
<b>Figura 18:</b> Diagrama cor (g-r) versus magnitude na banda i, em diferentes intervalos de <i>redshift</i> , dos zphotos obtidos no LePhare e zspecs do SVdeep, usando SEDs CFHTLS, com e sem distinção entre os tipos de galáxias Early e Late.....	44
<b>Figura 19:</b> Comparação das Magnitudes Absolutas estimadas com zphotos obtidos no LePhare com as obtidas com zspecs do VVDS, usando SEDs CFHTLS, com e sem a banda u .....	46
<b>Figura 20:</b> Esquema ilustrando o funcionamento do TPZ.....	48
<b>Figura 21:</b> Comparação de zphotos obtidos no TPZ com zspecs do VVDS sem a banda u.....	51
<b>Figura 22:</b> Comparação de zphotos obtidos no TPZ (4 florestas de 25 árvores) com zspecs do SVdeep .....	52
<b>Figura 23:</b> Comparação de zphotos obtidos no TPZ (3 florestas de 50 árvores) com zspecs do SVdeep .....	53
<b>Figura 24:</b> Comparação das Magnitudes Absolutas estimadas com zphotos obtidos no TPZ com as obtidas com zspecs do VVDS, usando SEDs CFHTLS, com e sem a banda u .....	54



## 1 Introdução

Muitos estudos nas áreas de evolução de galáxias, estrutura do Universo em grande escala e Cosmologia tiveram um impulso significativo a partir de mapeamentos de galáxias em grandes áreas do céu. Esses grandes levantamentos forneceram informações detalhadas sobre a estrutura do Universo, que por sua vez nos ajuda a compreender sua geometria e composição, assim como a formação e evolução das galáxias.

Entretanto, a partir dos anos 2000, o desafio para a pesquisa na área de Cosmologia atingiu outro patamar. A descoberta por Perlmutter et al. (1990) e Riess et al. (1990) de que a expansão do Universo é acelerada direcionou o grande interesse da comunidade científica para desvendar a causa ainda desconhecida da energia responsável por esta aceleração, denominada energia escura. Grandes projetos internacionais, liderados por instituições de primeira linha e grupos envolvendo centenas de pesquisadores, iniciaram mapeamentos de volumes do universo sem precedentes. De fato, as novas tecnologias envolvidas no desenvolvimento instrumental remetem a quantidades de dados a serem obtidos para o patamar de centenas de milhões a um bilhão de objetos a serem observados. Neste novo contexto de pesquisa, a obtenção de *redshifts*<sup>1</sup> espectroscópicos torna-se (pelo menos atualmente) um grande desafio. Assim, os mapeamentos fotométricos tornaram-se iniciativas de maior viabilidade para estudos em Cosmologia e Astronomia Extragaláctica, envolvendo

---

<sup>1</sup> O *redshift* é definido como o deslocamento para o vermelho do espectro de uma fonte que se afasta com uma velocidade de recessão  $v$  e medido como  $z = (\lambda_0 - \lambda_r)/\lambda_r = \frac{\sqrt{1+\frac{v}{c}}}{\sqrt{1-\frac{v}{c}}} - 1$ , em que  $\lambda_r$  é um determinado comprimento de onda do espectro da fonte em repouso,  $\lambda_0$  é um comprimento de onda observado e  $c$  é a velocidade da luz. Na ausência de movimentos peculiares, o *redshift* de uma galáxia pode ser associado a uma distância através de um modelo cosmológico representando a expansão do Universo.

grandes bases de dados. A expectativa (com fundamento em estudos preliminares) é de que as maiores incertezas na determinação de *redshifts* fotométricos (comparados aos espectroscópicos) sejam compensadas pela grande quantidade de dados.

Exemplos de mapeamentos visando à determinação de *redshifts* espectroscópicos podem ser citados como: CfA Redshifts Survey (Davis et al. 1982), Southern Sky Redshifts Survey (da Costa et al. 1988), 2-degree Field Redshifts Survey (2dF, Colless et al. 2001), WiggleZ Dark Energy Survey (Drinkwater et al. 2010), Barion Oscillation Spectroscopic Survey (BOSS, Dawson et al. 2013), e os futuros Dark Energy Spectroscopic Instrument (DESI, <http://desi.lbl.gov/>) e Euclid Consortium (<http://www.euclid-ec.org/>). Eles fornecem uma imagem tridimensional da distribuição de galáxias, mas são dispendiosos em tempo e recursos e podem sofrer de profundidade limitada e efeitos de seleção.

Exemplos de mapeamentos fotométricos são: Sloan Digital Sky Survey (SDSS, York D. G. et al. 2000), Canada-France Hawaii Telescope Legacy Survey (CFHTLS, Ilbert et al. 2006), Panoramic Survey Telescope and Rapid Response System (Pan-STARRS, Kaiser, Tonry & Luppino 2000), Dark Energy Survey (DES, Flaugher 2005, <http://www.darkenergysurvey.org/>) e o futuro Large Synoptic Survey Telescope (LSST, Tyson et al. 2003). Estes são mais eficientes na rapidez de obter dados e geralmente mais profundos e completos, mas não fornecem uma visão tridimensional precisa do Universo devido à sua limitada resolução nas posições das galáxias ao longo da linha de visada. Essas distâncias são estimadas através da medição do *redshift* fotométrico (photo-z) de cada galáxia usando-se os fluxos medidos de um conjunto de filtros de banda larga.

Em particular, o DES é um mapeamento fotométrico em operação que está produzindo uma quantidade de dados sem precedentes. No total serão mapeados 5000 graus quadrados do céu do sul em 5 bandas (g, r, i, z, Y), usando para isso a DECam, uma câmera de 570 mega-pixels construída pelo projeto que opera nas faixas do ótico e infravermelho próximo. Ela está situada no foco principal do telescópio Blanco (de 4m de diâmetro), do Cerro Tololo Interamerican Observatory (CTIO), no Chile. Uma estimativa confiável de *redshifts* fotométricos das galáxias é uma parte crucial desse programa, que espera determinar photo-zs com erro  $\sigma_{zs} < 0.12$ .

O grupo brasileiro envolvido no Dark Energy Survey atua através do Laboratório Interinstitucional de e-Astronomia (LIneA). Além de participar de diversos temas de ciência do projeto, a equipe tem como responsabilidade o desenvolvimento de seu portal científico, através do qual diversas amostras, análises comuns e projetos-chave serão realizados. Em particular, códigos de determinação de *redshifts* fotométricos estarão disponibilizados para a colaboração. Assim, a equipe dedicou esforços para o entendimento operacional de alguns códigos (bem como estimulou pesquisadores do projeto a participar) e para implementá-los no portal.

É dentro deste contexto do envolvimento brasileiro no DES que está escrita esta monografia.

Como a primeira avaliação da qualidade dos *redshifts* obtidos por uma variedade de códigos foi realizada por Sanchez et al (2014), utilizamos este trabalho como base para apresentar inicialmente, no capítulo 2, uma descrição sucinta de alguns desses métodos.

No capítulo 3 apresentamos resultados provenientes de dois tipos diferentes de métodos de determinações de photo-zs: um método de ajuste de *templates* (LePhare) e um método de treinamento (TPZ). O LePhare foi o código mais explorado pelo grupo brasileiro no DES, que contribuiu para Sanchez et al. (2014) com os resultados obtidos a partir dele, por sua capacidade de fornecer parâmetros característicos das galáxias (tais como tipo, magnitude absoluta, massa e idade). As avaliações do código de treinamento (TPZ) foram concentradas em amostras específicas com o objetivo de examinar possíveis melhorias nas determinações dos photo-zs.

Já no capítulo 4 apresentamos os comentários gerais sobre o trabalho e as conclusões.

Em algumas figuras apresentadas nesta monografia há termos em inglês, pois elas foram apresentadas em relatórios e documentos da colaboração internacional e reeditá-las demandaria um tempo apreciável, sem ganho real para a compreensão dos resultados aqui apresentados.

## 2 Métodos de determinação de photo-zs

### 2.1 Classificação dos métodos

Sanchez et al. (2014) realizaram uma avaliação dos dados do DES obtidos de novembro de 2012 a fevereiro de 2013. Estes foram denominados DES-SV por serem relativos a um período de observações de verificação científica (*science verification*, SV) do DES. Forneceram imagens de qualidade científica para cerca de 150 graus quadrados na profundidade nominal do mapeamento,  $z \sim 1.4$ . Os desempenhos dos diferentes métodos de photo-z foram avaliados através de comparações com informações espectroscópicas de outros mapeamentos. Para isso, cerca de 15 mil galáxias com *redshifts* espectroscópicos conhecidos foram usadas como amostra de referência para os valores verdadeiros de *redshift*.

Existem duas abordagens principais para a determinação de *redshifts* fotométricos:

**Métodos de Ajuste de *Templates***, onde a distribuição espectral de energia (SED) obtida dos fluxos é comparada a um conjunto de *templates* de galáxias, desviadas para o vermelho, até que uma melhor combinação seja encontrada, determinando assim tanto o tipo espectral e a extinção interna da galáxia, quanto o seu *redshift*. Um texto com uma descrição simplificada dos tipos de galáxias pode ser encontrado em <https://www.spacetelescope.org/images/heic9902o/>.

Exemplos: Hyperz (Bolzonella, Miralles & Pelló 2000), BPZ (Benitez 2000, Coe et al. 2006), LePhare (Arnouts et al. 2002, Ilbert et al. 2006), EAZY (Brammer, van Dokkum & Coppi 2008)

**Métodos de Treinamento**, onde um conjunto de galáxias com *redshifts* espectroscópicos conhecidos é usado para treinar um algoritmo, que é então aplicado sobre o conjunto de galáxias de interesse.

Exemplos: ANNz (Collister & Lahav 2004), ArborZ (Gerdes et al. 2010), TPZ (Carrasco Kind & Brunner 2013)

Cada técnica tem suas próprias vantagens e desvantagens, e uma combinação delas pode explorar esse fato, tirando benefícios das vantagens individuais de cada código.

De um modo sucinto, os principais resultados de Sanchez et al. (2014) mostraram que métodos empíricos de photo-z utilizando, por exemplo, Redes Neurais Artificiais ou Florestas Aleatórias, tiveram o melhor desempenho nos testes. No entanto, os resultados da maioria dos códigos, incluindo os métodos de ajustes de *templates*, atendem confortavelmente aos requisitos do DES no desempenho do photo-z.

Para que o photo-z seja útil em estudos cosmológicos, é preciso calibrá-lo, entendendo-se as propriedades estatísticas das diferenças entre os *redshifts* mais precisos (espectroscópicos) e as estimativas de photo-z: seu valor médio (para verificar tendências), sua dispersão (erros intrínsecos) e suas caudas (fração de photo-zs com grandes erros).

## 2.2 Descrição sucinta dos métodos

A descrição abaixo, baseada em Sanchez et al. (2014), não é completa, pois outros métodos podem ser encontrados na literatura e novos continuam a ser desenvolvidos. Entretanto, avaliamos que os códigos mencionados abaixo são os mais utilizados, usando como marco temporal o ano de 2014. Maiores detalhes sobre cada método podem ser encontrados na referências indicadas.

A maior parte desses códigos foi executada por grupos individuais, enquanto uma fração foi efetuada dentro do Portal de Ciência do DES elaborado pelo grupo brasileiro. Devido ao grande número de métodos utilizados por Sanchez et al. (2014), essa monografia, além de mostrar as capacidades do photo-z do DES-SV, também serve como uma referência útil para comparar diferentes códigos de photo-z usando dados reais (provenientes de observações) de um mapeamento de galáxias profundo.

Deve-se ressaltar que cada descrição resumida a seguir foi apresentada por grupos de pesquisadores que desenvolveram ou utilizaram exaustivamente o método, portanto suas sugestões de uso e de configurações são excelentes referências para novos usuários. Contudo, deve-se mencionar que essas escolhas são adequadas às bandas grizY e a alguns tipos de magnitude (no caso, MAGAUTO e MAGDETMODEL)<sup>2</sup> usadas no DES. É possível que diferentes conjuntos de bandas e tipos de magnitude possam produzir melhores resultados com diferentes escolhas de parâmetros dos códigos.

---

<sup>2</sup> MAGAUTO e MAGDETMODEL são tipos diferentes de determinações de magnitudes realizadas em imagens astronômicas pelo software SExtractor (Bertin e Arnouts, 1996), que identifica objetos e calcula diversos parâmetros. MAGAUTO se refere à magnitude determinada usando-se as contribuições de todos os pixels contíguos (acima de um nível escolhido de detecção) que formam o objeto. MAGDETMODEL se refere a uma magnitude determinada usando-se modelos de distribuição de luz ajustados ao conjunto de pixels que formam o objeto.

### 2.2.1 Dark Energy Survey Data Management (DESDM)

O DESDM foi computado usando o mesmo método de rede neural artificial aplicado à amostra do Sloan Digital Sky Survey Data Release 6 (DR6), descrito por Oyaizu H. et al. (2008). Em suma, foi utilizada uma rede neural com 10 nós de entrada, consistindo das 5 magnitudes grizY MAGAUTO e das 5 grizY MAGDETMODEL, seguidas por 3 camadas ocultas com 15 nós em cada. A minimização formal para determinar os pesos da rede neural foi feita sobre os dados do conjunto de treinamento. Já a escolha do conjunto de pesos da rede que produziu o menor espalhamento do photo-z no conjunto de testes foi feita depois de 300 iterações da minimização de peso. Além disso, para reduzir a sensibilidade às condições iniciais no procedimento de minimização, repetiu-se o procedimento 10 vezes, iniciando cada vez de uma posição inicial diferente no espaço de pesos. O photo-z final para uma galáxia foi considerado a média dos photo-zs calculados a partir dos pesos ideais para cada uma das 10 minimizações de rede.

Também calculou-se os erros de photo-z usando a técnica empírica do Nearest Neighbor Error (NNE), também descrita por Oyaizu H. et al. (2008). Esse método estima o erro do photo-z para cada galáxia empiricamente, com base nos photo-zs e *redshifts* espectroscópicos dos 100 vizinhos mais próximos do conjunto de testes espectroscópicos. Especificamente, o erro do photo-z do NNE é definido de forma que corresponda à largura de 68% da distribuição de  $|z_{phot} - z_{spec}|$  dos vizinhos mais próximos.



### 2.2.2 Artificial Neural Network (ANNz)

ANNz (Collister & Lahav 2004) é um método de treinamento que usa um esquema de rede neural para encontrar uma relação funcional entre um determinado conjunto de parâmetros de entrada (magnitudes, cores, etc.) e produz uma certa quantidade desejada (como o *redshift*). A arquitetura da rede neural é definida por  $c$  camadas e  $n$  nós, e treinada com um conjunto conhecido de parâmetros. Os nós estão interligados com pesos que são determinados minimizando-se uma função de custo usando os conjuntos de treinamento e testes com *redshifts* espectroscópicos conhecidos. Após muitos testes, encontrou-se o menor  $\sigma_{68}$  quando se utiliza uma arquitetura de 3 camadas com 10 nós em cada, e usando como entrada o conjunto de 5 magnitudes MAGAUTO e 5 MAGDETMODEL. Tentativas de se utilizar um maior número de nós e cores como entrada resultaram em erros de photo-z maiores. As incertezas na estimativa do photo-z são calculadas usando a propagação dos erros nas magnitudes de entrada para o erro no photo-z, usando a relação funcional entre estes parâmetros de entrada e o photo-z de saída.

### 2.2.3 Trees for Photo-z (TPZ)

TPZ (Carrasco Kind & Brunner 2013, 2014) é uma máquina de aprendizado (com processamento paralelo) que usa árvores de previsão e técnicas de florestas aleatórias para produzir tanto distribuições de probabilidades (*probability distribution function*, pdf) de *redshift* fotométrico robusto quanto informações auxiliares para uma amostra de galáxias. Uma árvore de previsão é construída fazendo-se uma sequência de perguntas que dividem os dados de entrada extraídos da amostra espectroscópica, frequentemente em dois ramos, até ser criada uma folha

terminal que atende a um critério de parada (como um tamanho mínimo de folha).

A dimensão em que os dados são divididos é escolhida para ser a única com maior ganho de informações em meio à subamostra aleatória de dimensões obtidas em cada ponto. Este processo produz árvores menos correlacionadas e permite a exploração de várias configurações dentro dos dados. A pequena região delimitadora dos dados no nó de folha terminal representa uma subamostra específica de todos os dados inteiros com propriedades semelhantes. Dentro dessa folha, é aplicado um modelo que proporciona uma previsão compreensível, especialmente em situações em que podem existir muitas variáveis que interagem de um modo não linear.

Usando os erros de magnitude dos dados, muitas árvores não correlacionadas podem ser criadas, cujos resultados são agregados para construir cada pdf individual. Para a aplicação aos dados do DES-SV, utilizou-se as magnitudes MAGAUTO e MAGDETMODEL nas 5 bandas do DES, juntamente com todas as cores correspondentes e seus erros associados.

#### **2.2.4 Relevance Vector Machine (RVMz)**

RVMz é um código de photo-z empírico com base no algoritmo de máquina vetorial desenvolvido por Tipping (2001), um método de regressão. Tal máquina vetorial, denominada Relevance Vector Machine (RVM), tem características semelhantes à máquina vetorial de suporte, mas inclui um tratamento Bayesiano para a determinação dos pesos do modelo. Isto tem a vantagem de que a RVM não requer validação para otimização de parâmetros. Para obter estimativas de photo-z, utiliza-se as magnitudes

MAGDETMODEL (grizY) e cores (g-r), (r-i), (i-z) e (z-Y) de entrada. Reconstrói-se o pdf combinando-se as incertezas nos conjuntos de dados e do modelo.

### 2.2.5 Normalized Inner Product – Nearest Neighbor (NIP-kNNz)

NIP-kNNz (Juan De Vicente et al. 2014) é uma técnica que computa o photo-z de uma aproximação NNE baseada no chamado Normalized Inner Product (NIP). A distância-magnitude euclidiana garante que galáxias próximas no espaço de magnitude possuem o mesmo *redshift*, mas não considera como vizinhas galáxias com a mesma cor, porém separadas em magnitude. A métrica NIP corrige isso considerando 2 galáxias como vizinhas, e portanto de *redshift* próximos, quando elas têm cores similares, ao invés das magnitudes. O NIP é baseado na seguinte definição:

$$NIP = \cos \alpha = \frac{\overrightarrow{M}_t \cdot \overrightarrow{M}_p}{M_t M_p},$$

em que  $\overrightarrow{M}_t$  e  $\overrightarrow{M}_p$  são os vetores de treinamento (*training vector*) e fotométrico (*photometric vector*), respectivamente, e  $\alpha$  é o ângulo entre eles. Para essa aplicação em particular, as 5 magnitudes MAGDETMODEL foram utilizadas e transformadas em fluxos. O NIP relaciona-se com o ângulo que os 2 vetores de multi-magnitude formam: maximizar NIP é equivalente a minimizar este ângulo.

### 2.2.6 Artificial Neural Network 2 (ANNz2)

ANNz2 é uma nova versão do ANNz que tornou-se pública em 2014. Este código incorpora métodos como redes neurais artificiais (ANNs), árvores de decisão (BDTs) e vizinhos mais próximos K (KNNs). Os diferentes algoritmos são utilizados para aperfeiçoar a reconstrução do desempenho do photo-z e estimar as incertezas das soluções fotométricas. Isso é feito gerando-se uma ampla seleção de soluções com diferentes arquiteturas ANN, inicializadas por distintas sementes aleatórias. Tal aperfeiçoamento é realizado enumerando-se diferentes soluções de acordo com seus desempenhos, que são determinadas pelo respectivo espalhamento de photo-z de cada solução.

A solução simples com a melhor performance é escolhida pelo estimador de photo-z do ANNz2. Além disso, o conjunto inteiro de soluções é usado para derivar a função de densidade de probabilidade (pdf) do photo-z, construída em 2 passos. Primeiramente, cada solução é guardada com um erro de distribuição, o qual é derivado usando-se o método de vizinhos próximos KNN de Oyaizu H. et al. (2008). O conjunto (*ensemble*) de soluções é então combinado, o que é feito pesando-se estimadores distintos, de forma a produzir uma pdf que descreva os erros fotométricos. As entradas usadas foram as 5 magnitudes MAGAUTO e as 5 MAGDETMODEL.

### 2.2.7 ArborZ

Algoritmo (Gerdes et al. 2010; Sypniewski 2014) estimador de photo-z baseado em um conjunto de treinamento de disponibilidade pública que utiliza árvores de decisão. As BDTs foram desenvolvidas para

classificar objetos caracterizados por um vetor de observáveis em 2 categorias. Uma árvore de decisão aleatória é um classificador relativamente fraco, mas a “floresta” de árvores de decisão gerada pelo processo de treinamento é um poderoso classificador. Para adaptar um classificador binário ao problema de determinar o *redshift*, divide-se a faixa de interesse deste último em  $N$  *bins* discretos com uma largura de cerca de 25-50% da resolução esperada do photo-z, e treina-se um BDT classificador para cada *bin* de *redshift* usando-se uma floresta de 50 árvores. Cada classificador é treinado para identificar as galáxias com *redshift* tendendo a seu *bin* particular e rejeitar as que tendem para mais do que  $3\sigma_{z_{phot}}$  do seu *bin* como “background”. Ao ser apresentado a uma nova galáxia, cada classificador BDT produz uma probabilidade de que a galáxia dada tenda em direção ao “*bin*” do seu *redshift*. O conjunto dessas probabilidades constitui a pdf do ArborZ, que por sua vez fornece a melhor estimativa de um photo-z individual (a partir da média da pdf) e seu erro (pela largura), mas também fornece a melhor caracterização do photo-z de uma galáxia como um todo.

Na amostra do DES-SV, treina-se o ArborZ usando-se as magnitudes MAGAUTO e MAGDETMODEL em grizY, com 50 *bins* de *redshift* de largura ajustada até  $z = 1.5$ . Os resultados são robustos em relação às variáveis no tamanho e número de *bins* da floresta.

### 2.2.8 Skynet

Este método, usado inicialmente em Bonnett (2013), consiste na utilização de uma rede neural para classificar galáxias em classes, neste caso, *bins* de *redshift*. Uma rede neural com uma saída *softmax* transformada (Pyle 1999) é capaz de estimar a probabilidade de que um

objeto pertença a uma determinada classe. Levando em conta que uma galáxia não pode permanecer em mais do que um *bin* de *redshift* ao mesmo tempo, uma rede neural com uma transformação *softmax* é ideal para retornar uma pdf para cada galáxia. Antes de se treinar a rede neural, é preciso dividir os dados em  $n$  *bins* de *redshift* (as classes). A rede neural é alimentada pelas magnitudes MAGAUTO e MAGDETMODEL e pelas classes corretas. A rede neural dá como saída  $n$  valores entre  $[0,1]$  para cada galáxia. Tais valores somam 1, portanto podem ser interpretados como a probabilidade de que uma galáxia resida em um *bin* de *redshift*. A rede neural é treinada usando-se 3 camadas ocultas com respectivamente 20, 30 e 40 nós em cada.

### 2.2.9 Bayesian Photometric Redshift (BPZ)

O BPZ, código de photo-z de Benitez (2000) e Coe et al. (2006), é um método baseado em *templates* que retorna a distribuição de densidade de probabilidade  $p(z|m_i)$  que a galáxia está em *redshift*  $z$ , quando suas magnitudes em cada banda são  $m_i$  e também um único valor de photo-z é calculado como o máximo de  $p(z|m_i)$ . De acordo com o Teorema de Bayes,  $p(z|m_i)$  é o produto entre uma probabilidade e uma função de probabilidade *prior* que representa nosso conhecimento anterior do *redshift* e de distribuições do tipo espectral da amostra em análise. São usadas as 5 magnitudes MAGDETMODEL.

*Templates*: São utilizadas as oito *templates* espectrais que o BPZ carrega pelo padrão baseado em Coleman, Wu & Weedman (1980); Kinney et al. (1996), e são adicionadas duas *templates* a mais, interpoladas entre cada par deles, definindo-se o parâmetro de entrada INTERP = 2 (opção padrão).

*Prior* (resultado previamente conhecido para a distribuição de galáxias em função do *redshift*, do tipo e da magnitude): Calibra-se o *prior* em cada teste encaixando-se a função empírica  $\Pi(z, t|m_0)$  proposto em Benitez (2000) para a correspondente amostra de treinamento do DES-SV. Com isso, pode-se remover a maior parte dos *outliers* catastróficos que, para métodos baseados em *templates*, tendem a constituir uma fração considerável de todas as galáxias.

### 2.2.10 Photometric Analysis for Redshift Estimations (LePhare)

LePhare (Arnouts et al. 2002, Ilbert et al. 2006) é um código de ajuste de *templates* que usa a minimização de diferenças entre as magnitudes teóricas e as observadas para encontrar a melhor *template* e o melhor *redshift*. Fornece para cada objeto uma estimativa de máxima probabilidade do *redshift*, correções K, magnitudes absolutas para cada banda e uma função de distribuição de probabilidade. Um método adaptativo e um *prior*  $N(z)$  podem ser usados para melhorar as magnitudes teóricas e minimizar os erros catastróficos, respectivamente. Diversos conjuntos de SEDs e leis de extinção estão disponíveis no código para uso. Para esse estudo utilizam-se as 5 magnitudes MAGDETMODEL.

*Templates*: Um conjunto de 66 SEDs foi usado na análise dos dados Canada-France Hawaii Telescope Legacy Survey (CFHTLS) por Ilbert et al. (2006), tendo estas sido obtidas por interpolação das *templates* amplamente usadas e publicadas por Coleman, Wu & Weedman (1980) para diferentes tipos de Hubble e as publicadas por Kinney et al. (1996) para galáxias *starbusts*. Uma vez que códigos de ajuste de *templates* consomem muito tempo, buscou-se dentro deste grande conjunto um grupo

reduzido de *templates* que possuísse essencialmente as mesmas características totais do agrupamento maior, considerando-se dispersão e fração de *outliers*. Ou seja, desejava-se um conjunto que representasse bem o total. Foram desenvolvidos diversos testes usando a amostra do VIMOS VLT Deep Survey (VVDS, LeFevre et al. 2014) com *redshifts* espectroscópicos disponíveis e conseguiu-se bons resultados com um grupo reduzido de 21 *templates* abrangendo SEDs de 12 galáxias elípticas, 6 espirais, 2 *starbusts* e uma irregular. Muitos testes realizados pelo grupo brasileiro removendo a banda u dos dados de VVDS e CFHTLS indicaram que as discrepâncias do photo-z do valor real (isto é, espectroscópico) aumentam devido à degenerescência galáctica. Concluiu-se que para minimizar esse problema era necessário manter os valores de extinção  $E(B-V)$  menores ou iguais a 0,25 (para os tipos tardios e Scd) e utilizar apenas 3 tipos muito tardios de SEDs (uma irregular e 2 *starbusts*).

*Prior*: Embora o *prior*  $N(z)$  disponível refira-se à cor (B-I) do mapeamento VVDS, verificou-se que aplicando o procedimento para a cor (g-i) dos dados do DES se obteria uma minimização significativa da fração de *outliers* (devido a grandes erros de photo-z), de forma que esta última acabou sendo a escolhida.

Treinamento: O método de treinamento disponível foi usado para obter uma recalibração em cada banda. Esse procedimento foi inicialmente aplicado aos conjuntos de treinamento e os deslocamentos obtidos foram usados quando colocou-se amostras para serem executadas no código.



### 2.2.11 Zurich Extra-galactic Bayesian Redshift Analyser (ZEBRA)

O ZEBRA de Feldmann et al. (2006) é um código de *redshift* fotométrico flexível e poderoso baseado no ajuste de *templates*. Produz uma distribuição posterior para cada galáxia no espaço de *redshift* e *templates*,  $P(z,T)$ , assim como distribuições marginalizadas para  $P(z)$  e  $p(T)$ . Para computação de *redshift*, filtros e *templates* foram suavizados sobre uma escala de 100 e 30 Angstroms, respectivamente. Nesta análise utilizou-se as 5 magnitudes MAGDETMODEL em grizY.

*Templates*: 81 *templates* foram usadas, tendo elas sido selecionadas de um super conjunto de SEDs. São as que obtiveram melhor ajuste a  $z = z_{\text{spec}}$  para galáxias da amostra de treinamento. Tal super conjunto foi produzido pelo módulo de correção de *templates* do ZEBRA a partir de interpolações logarítmicas entre as *templates* de Coleman, Wu & Weedman (1980) e Kinney et al. (1996). Esse módulo permite que o usuário defina intervalos de *redshifts* dentro daqueles em que as *templates* foram modificadas para melhor se ajustarem à fotometria de entrada. Para tal, uma amostra fotométrica do campo do mapeamento COSMOS (Scoville et al. 2007) foi usada. Uma fração das galáxias do conjunto de teste possui contrapartidas dentro desta amostra fotométrica, mas a fotometria é independente dos dados do DES e os *redshifts* espectroscópicos não foram usados.

*Prior*: No modo Bayesiano, o código ZEBRA constrói um *prior* iterativo auto-consistente usando funções de probabilidade de galáxias,  $L(z,T)$ . A aproximação é parecida com a de Brodwin et al. (2006), mas opera em um espaço bidimensional de *templates* e *redshifts*. O antecedente

construído a partir da amostra de treinamento foi usado em ambas as amostras.

Treinamento: Bordoloi, Lilly & Amara (2010) descrevem um método de usar *redshifts* espectroscópicos conhecidos para corrigir distribuições de probabilidade individuais de *redshifts* marginalizados,  $P(z)$ . Tal método demanda que a amostra de *redshifts* espectroscópicos e sua distribuição de probabilidades cumulativas entre zero e  $z_{\text{spec}}$  sejam planos. Aplicou-se um primeiro passo de suas aproximações em contribuições de *redshift*, com largura  $\Delta z_{\text{phot}} = 0.1$ . Galáxias foram atribuídas a esses *bins* baseadas em seus *redshifts* posteriores máximos. Após a correção dos  $P(z)$  individuais, um novo  $z_{\text{phot}}$  foi computado como o máximo dos  $P(z)$  corrigidos.

### 2.2.12 PhotoZ

PhotoZ (Bender et al. 2001) é um código fotométrico de ajuste de *templates* no qual a probabilidade do *redshift* de um objeto é dada multiplicando-se a probabilidade de ajuste de SEDs por probabilidades *priors* de *redshift* e luminosidade. A probabilidade total de um modelo é então:

$$P(\vec{\mu}|m) \propto \mathcal{L}(m|\vec{\mu}) \cdot P(\vec{\mu}),$$

em que  $m$  representa os dados fotométricos e  $\vec{\mu}$  são os parâmetros do modelo, isto é, *redshift*  $z$  e magnitude absoluta  $M$ . Nesta análise utilizou-se as 5 magnitudes MAGDETMODEL em grizY.

*Templates*: Este conjunto contém *templates* variando da faixa de formação estelar (azuis) até a de galáxias evoluindo passivamente (vermelhas). Isso inclui SEDs dos modelos de Bender et al. (2001), que foram criadas a partir de objetos espectroscopicamente observados pelo Hubble Deep Field North. Outras três *templates* (uma SO, uma Sab e uma galáxia Sbc) vêm de Mannucci et al. (2001) e duas SEDs empíricas vêm de Coleman, Wu & Weedman (1980). A esse conjunto modelo foram adicionadas ainda 13 SEDs de Ilbert et al. (2006), as quais foram otimizadas para combinar com galáxias locais de formação estelar. Essa é uma combinação de conjuntos de *templates* já usadas no passado para estimativa de *redshifts* fotométricos (Bender et al. 2001 e Brimiouille et al. 2013). Foi incorporado ainda neste conjunto um grupo de SEDs vermelhas para que se combinasse com as cores de galáxias vermelhas luminosas (Luminous Red Galaxies, LRGs) do SDSS-II (Greisel et al. 2013).

*Prior*: Os *priors* de *redshift* e luminosidade têm forma  $P(x) \propto \exp(-(x - X)/\sigma)^p$ , sendo  $x$ ,  $X$  e  $\sigma$  definidos individualmente para cada modelo de SED. Manuseando estas grandezas, é possível diminuir a probabilidade de se observar galáxias muito brilhantes em magnitude absoluta e modelos vermelhos a *redshifts* maiores ( $z > 0.9$ ). Além disso, pode-se adaptar os *priors* de  $z$  e  $M$  para cada SED de uma forma que *outliers* (erros catastróficos) de *redshift* fotométrico com  $|z_{phot} - z_{spec}| \div (1 + z) > 0.15$  nos conjuntos de treinamento das Verificações Científicas do DES sejam menos prováveis.

Treinamento: Os pontos zero dos catálogos de treinamento foram iterativamente adaptados usando-se as magnitudes médias entre os dados e as previsões do modelo ao se otimizar o desempenho do *redshift* fotométrico.

### 3 Avaliação da performance de códigos de photo-zs

Os resultados mostrados nesta seção representam meu projeto de Iniciação Científica, no qual aprendi, acompanhei e realizei, em paralelo ao trabalho do grupo brasileiro, uma parte dos mais básicos testes de determinação de *redshifts* fotométricos utilizando dois tipos de métodos: um de ajuste de *templates* (LePhare) e outro de árvores de decisão (TPZ). A maioria destes testes foram realizados com o objetivo de compreender e avaliar o funcionamento destes códigos, bem como suas limitações e vantagens, dando uma contribuição para os resultados publicados em Sanchez et al. (2014). Em particular, as principais questões a serem examinadas neste período foram a escolha dos métodos mais eficientes para a determinação dos photo-zs, o impacto causado pela ausência da banda u nos dados do DES e as primeiras estimativas das incertezas na determinação de magnitudes absolutas provenientes dos erros nas determinações de photo-z.

A avaliação dos resultados dos códigos neste trabalho se pautou nos requisitos determinados pela colaboração DES, que visam obter a equação de estado para a energia escura com erro compatível com os experimentos de estágio III do Dark Energy Task Force (Albrecht et al. 2006). Para atingir esse objetivo, a precisão dos resultados obtidos dos observáveis do projeto (contagem de aglomerados em função do *redshift*, escalas de aglomeração de galáxias, distâncias de supernovas e efeito fraco de lenteamento) será alcançada com os seguintes requisitos nas métricas de *redshift*:

a) Incerteza no photo-z (desvio padrão  $\sigma_{\text{es}}$  da relação photo-z *versus* *redshift* espectroscópico) inferior a 0.12;

b) Fração de objetos com erros estimados acima de  $2\sigma$  ( $f > 2\sigma$ ) inferior a 0.10;

c) Fração de objetos com erros estimados acima de  $3\sigma$  ( $f > 3\sigma$ ) inferior a 0.015.

Três amostras com *redshifts* espectroscópicos foram utilizadas nestes testes e incluem:

- O VVDS (LeFevre et al. 2014), que contém fotometria UBVRI e *redshifts* espectroscópicos de boa qualidade e profundidade de  $z = 1.4$ , semelhante a do DES. A amostra do campo F14 tem área comum com o CFHTLS campo Deep1, que tem fotometria ugriz, sistema similar ao DES. Assim, para realizarmos comparações e analogias com os dados do DES, utilizamos uma amostra com as galáxias do campo VVDS-F14 com seus *redshifts* espectroscópicos e fotometria ugriz, que passa a ser referida neste texto como VVDS. Esta amostra contém 3028 galáxias numa área de céu de cerca de 1 grau quadrado, então devido à pequena área do seu campo, sofre de variância cósmica. Entretanto, sua qualidade fotométrica, homogeneidade e representatividade de diferentes tipos de galáxias a tornaram muito utilizada para referência de *redshifts* “verdadeiros” e constitui uma das amostras de referência de Sanchez et al. (2014), já que esse campo também tem área comum com o DES. A magnitude limite desta amostra na banda  $i$  é 24.0 mag.
- Uma amostra simulada procurando reproduzir os futuros dados observacionais do projeto. Esta foi produzida para um desafio interno no projeto de que diferentes grupos reproduzissem, a partir

de seus códigos, os parâmetros cosmológicos utilizados para a geração do catálogo simulado (só conhecido pelos criadores da simulação). Utilizamos para alguns testes uma pequena parte da versão 6B dessa amostra, denominada Data Challenge 6B (DC6B).

- A amostra de Verificação Científica do DES descrita em Sanchez et al. (2014), que possui a profundidade esperada para o mapeamento e cujo subconjunto *deep* (SVdeep) inclui 15607 galáxias. A magnitude limite desta amostra na banda *i* é 23.5 mag. Para o caso de códigos que necessitam de amostras de treinamento e validação, o Grupo de Trabalho de *Redshifts* Fotométricos da colaboração DES definiu (para testes comuns) uma subdivisão desta amostra em 7249 galáxias para treinamento e 8358 para validação, o que satisfaz um dos requisitos do projeto para suas análises finais, que é utilizar dados com erros fotométricos inferiores a 0.01 mag.

A banda *i* foi escolhida como a de referência para magnitudes ao longo dessa monografia e a cosmologia utilizada foi  $(H_0, \Omega_m, \Lambda) = (70, 0.3, 0.7)$ .

### 3.1 LePhare

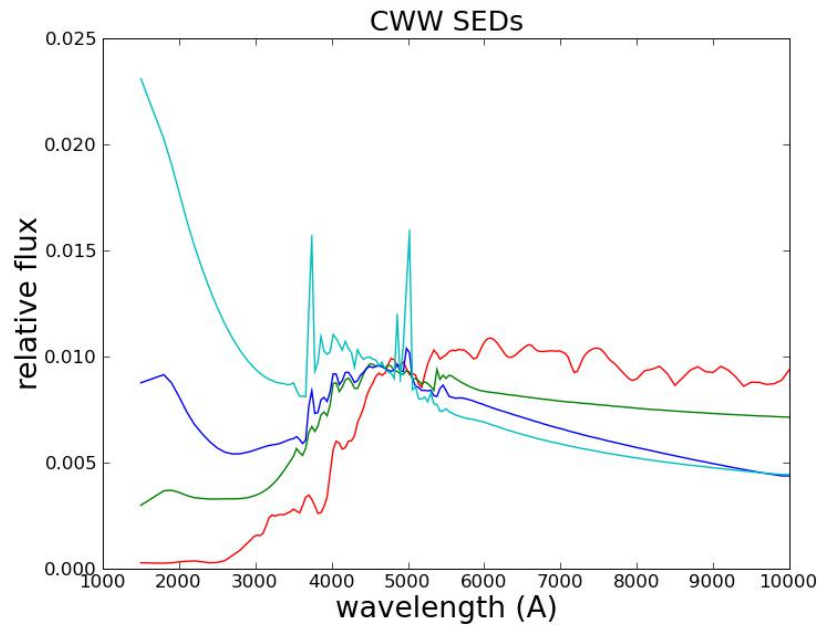
Como mencionado anteriormente, o LePhare é um código de ajuste de *templates* que usa a minimização de diferenças entre as magnitudes observadas e as teóricas obtidas da convolução de distribuições espectrais de energia (SEDs), deslocadas em *redshift*, com os filtros do mapeamento. O melhor ajuste indica qual SED, deslocada por um dado valor de photo-z, apresenta a melhor representação das magnitudes observadas.

O código é público, em linguagem Fortran, é de fácil instalação e inclui ferramentas de visualização dos resultados. Suas etapas (definidas num arquivo de configuração) são:

- a) Escolha do conjunto de SEDs;
- b) Escolha do conjunto de filtros;
- c) Escolha de *bins* e limites de *redshift*, escolha de limites e lei de extinção e SEDs passíveis de extinção, parâmetros cosmológicos e criação da biblioteca teórica;
- d) Escolha de uso de *prior*  $N(z)$  para redução de erros catastróficos e cálculo dos parâmetros de saída (photo-z, correções-K, magnitudes absolutas, etc...).

Muitos códigos de ajuste de *templates* utilizam o conjunto observado apresentado em Coleman, Wu e Weedman (1980, CWW), incluindo quatro espectros para os tipos Elíptica, Sbc, Scd e Im, algumas vezes complementado pelos espectros de galáxias *starburst* de Kinney et al.

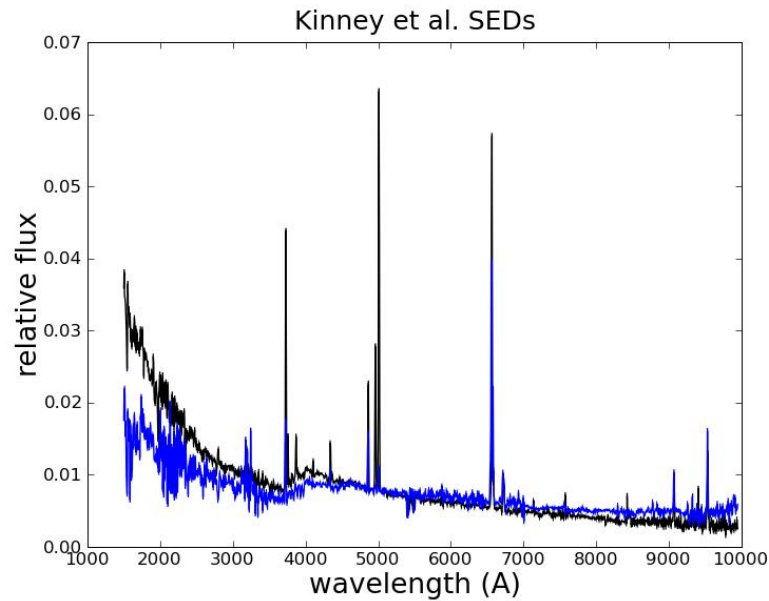
(1996). Eles representam um conjunto padrão utilizado em muitas análises de amostras de galáxias. A Figura 1 mostra o conjunto de SEDs CWW.



**Figura 1:** SEDs CWW. Elípticas em vermelho, Sbc em verde, Scd em azul e Im em ciano.

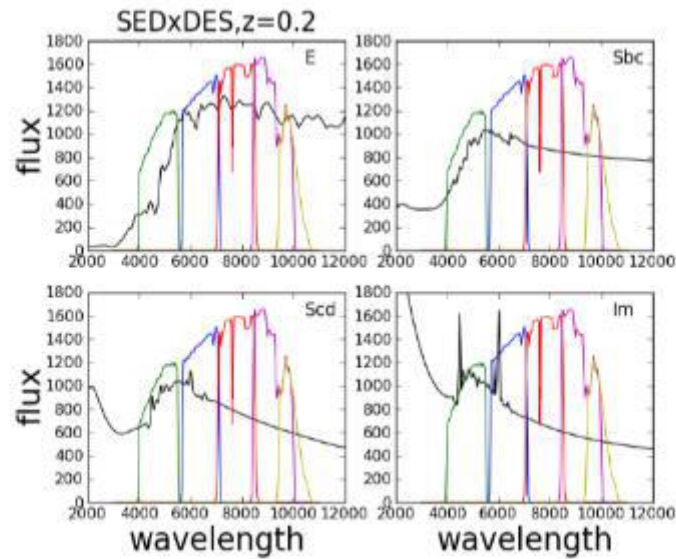
Na Figura 2 são mostradas duas SEDs do conjunto observado por Kinney et al. (1996). Em ambas as figuras os fluxos relativos estão normalizados em 4500 Angstroms.



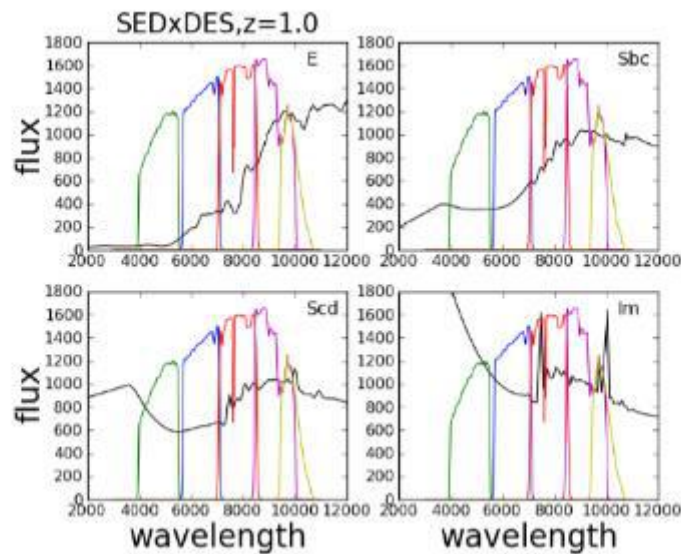


**Figura 2:** Duas SEDs (SB1 e SB3) do conjunto de *starbursts* de Kinney et al. Tipo SB1 em preto e SB3 em azul.

A essência do código LePhare está na criação de uma biblioteca de magnitudes teóricas obtidas pelo deslocamento das SEDs em *redshift* e multiplicando-as pelas transmissões dos filtros do mapeamento. Nas Figuras 3 e 4 abaixo são mostradas como exemplos, respectivamente, as distribuições espectrais de diferentes SEDs deslocadas em *redshift* = 0.2 e em *redshift* = 1.0, superpostas aos filtros do DES. Figuras como estas ajudam na compreensão de como as principais assinaturas dos espectros de galáxias causam (ou não) diferenças significativas nas magnitudes em bandas distintas em diferentes *redshifts*, o que é crítico para a determinação confiável dos photo-zs.



**Figura 3:** SEDs de galáxias Elípticas (E), Sbc, Scd e Im deslocadas em  $redshift = 0.2$ , e os filtros grizY do DES.

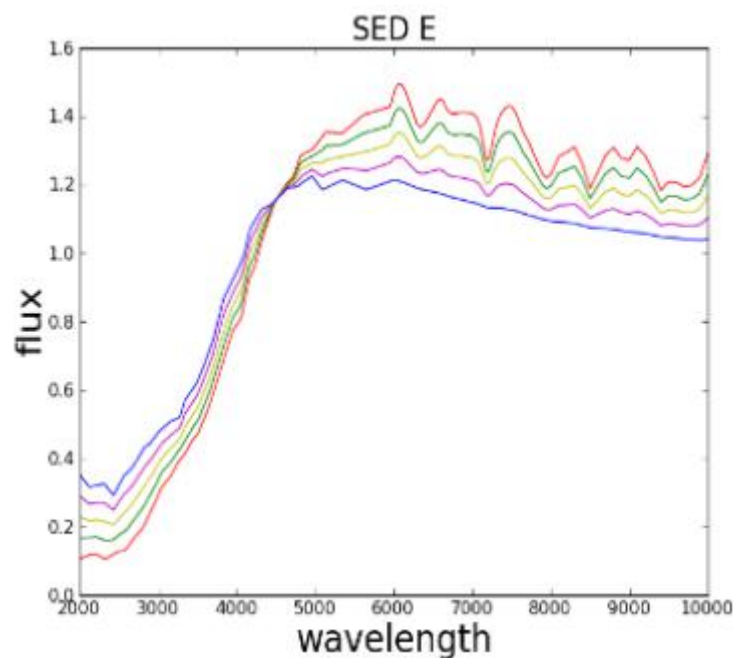


**Figura 4:** SEDs de galáxias Elípticas (E), Sbc, Scd e Im deslocadas em  $redshift = 1.0$ , e os filtros grizY do DES.

As SEDs CWW-Kinney representam, entretanto, um conjunto limitado de *templates*, o que pode prejudicar na determinação mais precisa dos photo-zs (como será visto mais adiante) ou não contemplar tipos intermediários de espectros de galáxias. Para contornar esse problema, o CFHT Legacy Survey elaborou para o seu mapeamento um conjunto de SEDs descritos em Ilbert et al. (2006) consistindo da interpolação de 10

*templates* entre cada uma das SEDs CWW e 2 das SEDs Kinney, totalizando 66 *templates*.

Em virtude da enorme quantidade de dados do DES e do tempo de computação necessário para rodar o código LePhare com este grande número de SEDs, o grupo brasileiro optou por examinar a possibilidade de reduzir o conjunto de SEDs CFHTLS sem perda da qualidade dos photo-zs obtidos. Isto foi possível com um subconjunto de 21 SEDs, sendo 12 de galáxias Elípticas, 6 de Espirais, uma de Irregular (Im) e 2 de *starbursts*. A Figura 5 mostra alguns tipos intermediários de SEDs de galáxias Elípticas, onde pode-se perceber a gradativa relevância de assinaturas espectrais (como bandas de absorção e linhas em emissão) que as SEDs interpoladas possibilitam considerar no cálculo de photo-zs.



**Figura 5:** SEDs de subtipos de galáxias Elípticas obtidos no CFHTLS por interpolação.

Deve-se mencionar que diversos outros conjuntos de SEDs estão disponíveis com o código LePhare, que incluem *templates* observadas ou

teóricas utilizadas em diferentes projetos (além do CFHTLS), como o Cosmos Survey. Além disso, estão disponibilizadas SEDs sintéticas conhecidas e largamente utilizadas na literatura como o PEGASE2 (Fioc e Rocca-Volmerange, 2000) e Bruzual e Charlot (2003), que podem fornecer estimativas de idade, massa e taxa de formação estelar, sendo de grande valor para estudos de formação e evolução de galáxias. Novos conjuntos de SEDS podem também ser incluídos ou combinados.

A biblioteca criada pelo LePhare contempla a possibilidade de acrescentar extinção às magnitudes teóricas utilizando as leis de Prevot et al. (1984) e Calzetti et al. (2000). As duas apresentam diferenças importantes, especialmente no ultravioleta. Entretanto como os dados atuais do DES não incluem a banda u, praticamente não existem diferenças quando se utiliza uma ou a outra lei. Nos dados apresentados nesta monografia, a lei de Prevot (1984) foi utilizada nas execuções do código LePhare.

Utilizamos também um recurso do código que usa uma distribuição esperada de *redshifts* para cada tipo de galáxia e sua magnitude observada para identificar photo-zs muito diferentes desse valor esperado. Estes têm grande probabilidade de possuírem erros enormes (erros catastróficos) e portanto são substituídos pelos valores esperados. Essas distribuições (*priors*) estão previamente definidas e disponibilizadas no código.

Em síntese, o LePhare utiliza a biblioteca por ele criada para buscar o *redshift* que minimiza a diferença entre as magnitudes observadas e as teóricas (dessa biblioteca), identificando assim um tipo (uma SED), um

valor de extinção e a correção-K<sup>3</sup> associada. Com esses dados, a magnitude absoluta da galáxia é determinada.

Não consideramos correções das SEDs devido a efeitos de linhas de emissão nas magnitudes teóricas, já que as *templates* utilizadas são espectros observados e já contêm essas linhas.

Nesta monografia utilizamos para as amostras do DES as magnitudes MAGAUTO usuais fornecidas pela aplicação do código SExtractor, disponibilizadas nos dados das amostras aqui utilizadas.

### 3.1.1 Photo-zs determinados com o LePhare

#### 3.1.1.1 Padrão de referência

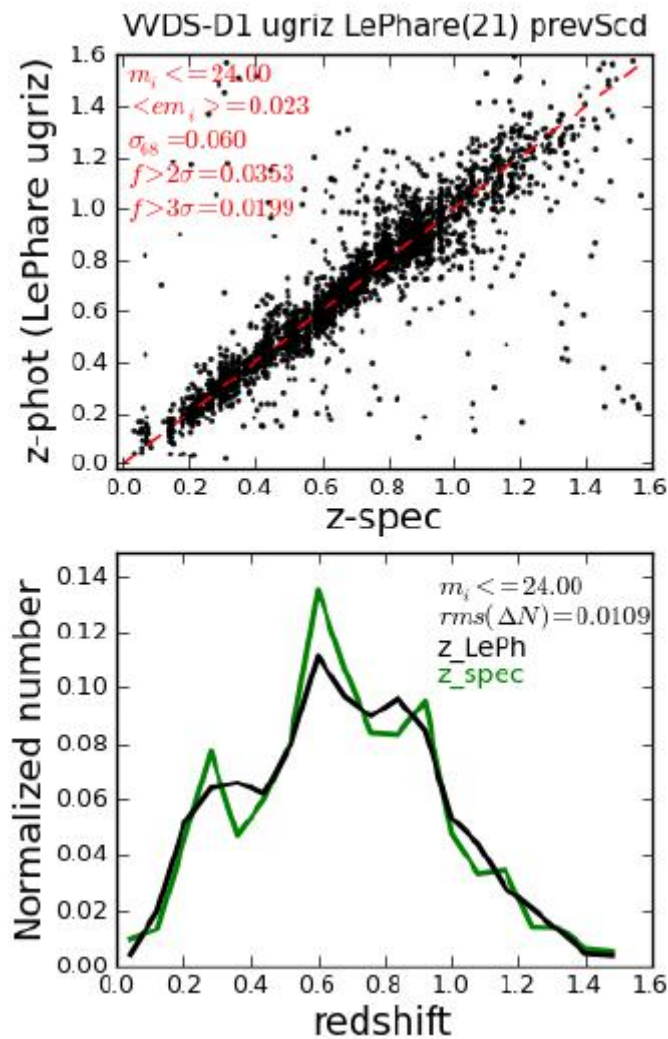
Para definir um padrão caracterizando um bom resultado de photo-z obtido com 5 filtros de banda larga, tomamos como referência a amostra VVDS (mencionada no início da seção 3) que inclui a banda u, de reconhecida importância para a determinação de *redshifts* fotométricos. Estes dados (embora poucos) constituem uma boa referência de uma amostra profunda com fotometria em 5 bandas (ugriz) e espectroscopia de boa qualidade, com a qual resultados de photo-zs podem ser obtidos e testados.

A Figura 6 mostra os resultados da comparação entre os photo-zs obtidos com o código LePhare e os *redshifts* espectroscópicos para o VVDS. Os valores indicados no painel superior da figura representam a

---

<sup>3</sup> A correção-K representa a diferença na determinação de uma magnitude de uma galáxia em seu referencial de repouso com aquela medida para a mesma fonte em um determinado *redshift*, ou seja em movimento com relação a Terra (expansão de Hubble, movimento peculiar).

magnitude da amostra na banda  $i$  ( $m_i$ ), a média dos valores dos erros em  $m_i$ ,  $\langle em_i \rangle$ , o valor do desvio padrão  $\sigma_{es}$  (ao nível de 68%), e as frações de objetos com erros acima de  $2\sigma$  ( $f > 2\sigma$ ) e  $3\sigma$  ( $f > 3\sigma$ ). Os resultados de photo-z foram obtidos utilizando-se o conjunto reduzido de 21 SEDs, aplicando-se 6 valores de extinção, relativos a excessos de cor desde  $E(B-V) = 0.0$  (nenhuma extinção) até  $E(B-V) = 0.25$ , para tipos Scd e mais *late*.



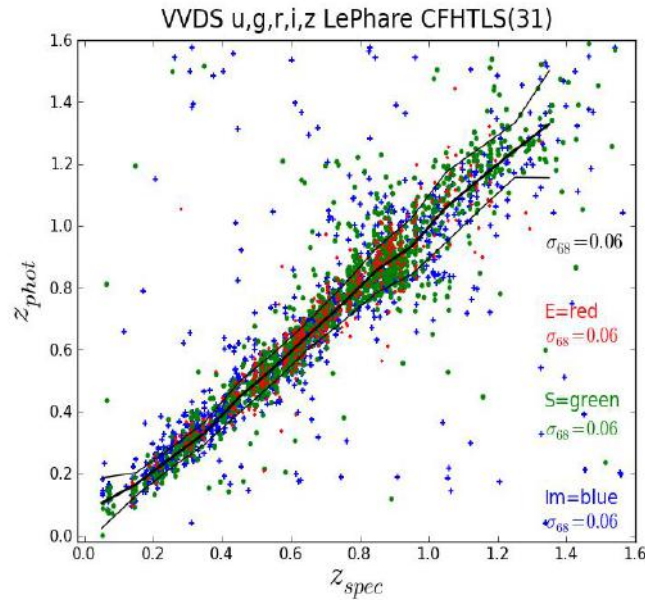
**Figura 6:** Comparação de resultados de photo-zs obtidos com o código LePhare e *redshifts* espectroscópicas da amostra VVDS com a banda u (painel superior). Comparação das distribuições destes dois tipos de *redshifts* (painel inferior).

O desvio padrão médio  $\sigma_{68} = 0.06$  é bastante satisfatório, uma vez que o requerimento científico do DES é de um valor inferior a 0.12. Dessa forma, uma fração grande das galáxias está dentro do erro de 68%, conforme o desejado. O valor da fração de galáxias superior a dois desvios padrões  $f > 2\sigma = 0.035$  está também significativamente abaixo do limite desejado para o DES, que é inferior a 0.10. Apenas a fração de galáxias  $f > 3\sigma = 0.02$  encontra-se ligeiramente acima do valor limite do DES, que é 0.015. Além disso, pode-se perceber que o LePhare produz *redshifts* fotométricos com viés bastante baixo, como se pode ver pela linearidade da distribuição dos pontos, comparando-se com a linha vermelha tracejada que representa a relação 1:1.

Um ponto importante para o projeto, especialmente para suas análises da distribuição de galáxias em grande escala, é que a distribuição de *redshifts*  $N(z)$ , representando o número de galáxias em *bins* de *redshifts*, seja representativa da distribuição real. No painel inferior da Figura 6 está apresentado de forma contínua o histograma normalizado de photo-zs, em preto, comparado com o de *redshifts* espectrocópicos, em verde. O valor  $\text{rms}(\Delta N)$  é uma das estatísticas de avaliação da eficiência em reproduzir *redshifts* espectrocópicos. Ele é o desvio quadrático médio das diferenças  $(z_{\text{phot}} - z_{\text{spec}})$  calculado para todos os *bins* do histograma normalizado. Considerando-se as incertezas intrínsecas envolvidas na determinação de photo-zs, pode-se avaliar como bastante satisfatória a reprodução da distribuição dos *redshifts* verdadeiros pelo código LePhare.

É interessante verificar como os diferentes tipos de galáxias contribuem para a dispersão dos resultados. Para tal, a Figura 7 mostra a mesma comparação de *redshifts* da Figura 6, painel superior, com as galáxias separadas em tipos atribuídos pelo LePhare. As Elípticas estão

representadas em vermelho, as Espirais em verde e as Irregulares em azul. Os desvios padrões idênticos para todos os tipos de galáxias significam que o código funciona com igual precisão para os três.



**Figura 7:** *Redshifts* fotométricos ( $z_{phot}$ , LePhare) versus *redshifts* espectroscópicas ( $z_{spec}$ , VVDS). Galáxias Elípticas representadas em vermelho, Sbc e Scd em verde e Irregulares Im em azul.

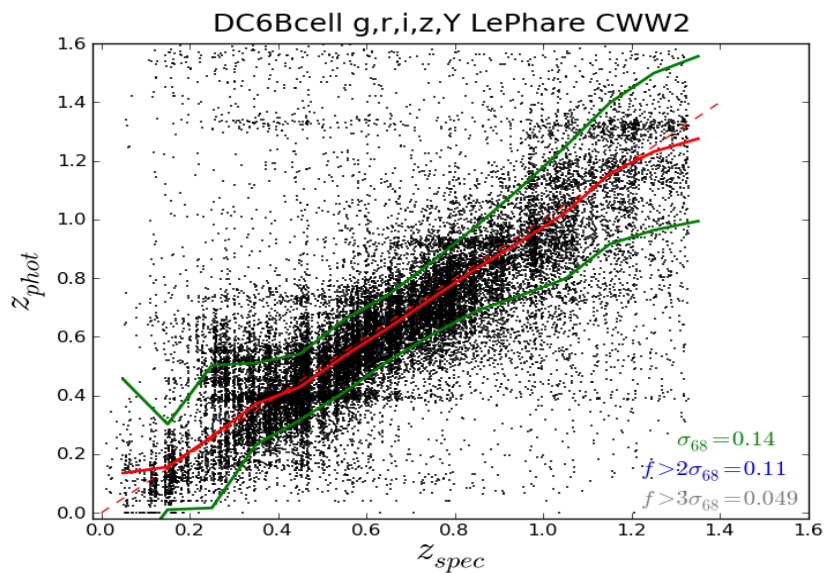
### 3.1.1.2. Amostra simulada do DES (DC6B)

Desde seu início, o DES também vem produzindo amostras simuladas de dados observacionais futuros. Uma dessas amostras foi elaborada para um desafio, dentro da colaboração, mencionado no início desta seção 3. Aplicamos o código LePhare numa subamostra de uma dessas versões de simulação, denominada DC6B. Do total de centenas de milhões de galáxias, selecionamos um conjunto de somente 30 mil objetos (uma célula do DC6B), tanto para reduzir o tempo de execução dos testes quanto pela finalidade, que foi apenas a de avaliar a qualidade dos photo-zs

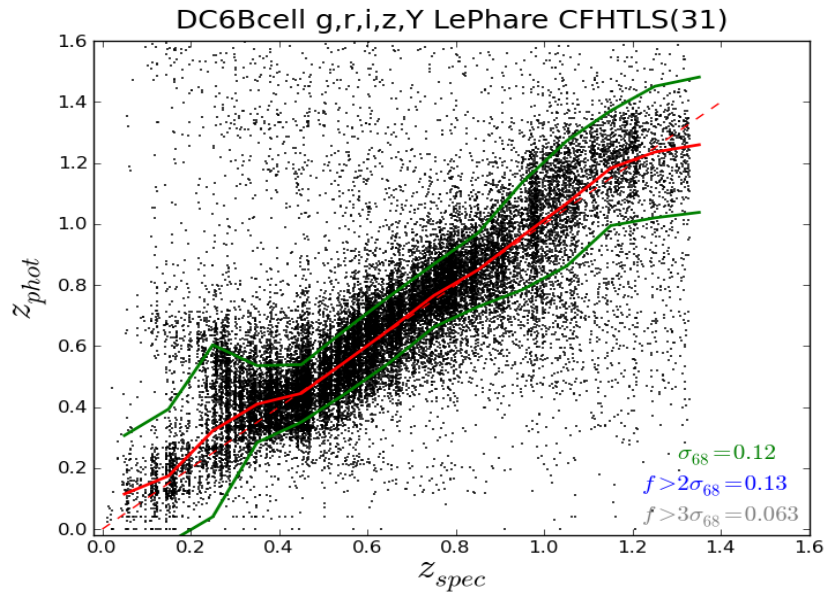


obtidos. É importante ressaltar que, por simular dados característicos do DES (g,r,i,z,Y), esta amostra não inclui magnitudes na banda u.

Primeiramente verificamos as diferenças causadas pela escolha de dois conjuntos de SEDs diferentes, principalmente no número de *templates*. Para tal, comparamos os resultados provenientes de um conjunto simples como o de 6 SEDs CWW-Kinney e um conjunto de 21 SEDs (reduzido do total de 66) do CFHTLS. Os resultados da comparação dos *redshifts* fotométricos obtidos com o LePhare com os *redshifts* produzidos na simulação são mostrados nas Figuras 8 e 9.



**Figura 8:** Comparação dos *redshifts* fotométricos obtidos pelo LePhare com os *redshifts* verdadeiros de uma simulação do projeto DES, utilizando o conjunto de SEDs CWW.



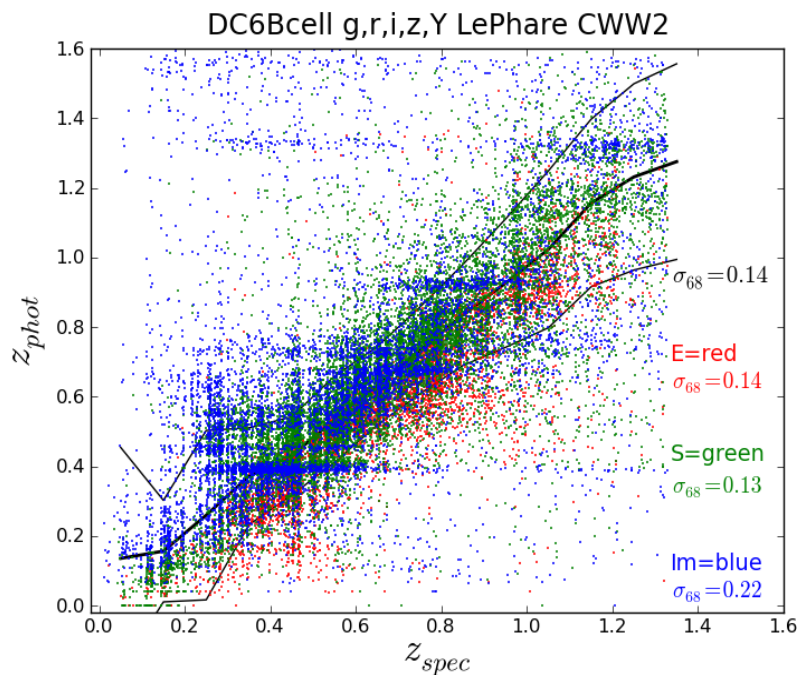
**Figura 9:** Comparação dos *redshifts* fotométricos obtidos pelo LePhare com os *redshifts* verdadeiros de uma simulação do projeto DES, utilizando o conjunto de SEDs CFHTLS.

Em primeiro lugar, nessas figuras pode ser constatado que o desvio padrão cresce significativamente (para ambos os conjuntos de SEDs) quando comparado com os resultados da análise da amostra VVDS, assim como um desvio localizado é claro para *redshifts* espectroscópicos entre 0.2 e 0.4, os quais são levados para valores de *redshifts* fotométricos entre 0.4 e 0.6. Este efeito é consequência da ausência da banda u nos dados do DES. Com isso, pode-se verificar que, no caso dos dados simulados, o desvio padrão médio está acima dos  $\sigma_{68} = 0.12$  desejado pelo projeto para as 6 SEDs CWW-Kinney. Verifica-se também a concentração de photo-zs nos *bins* construídos pela biblioteca do código, o que fica ressaltado com um baixo número de *templates*.

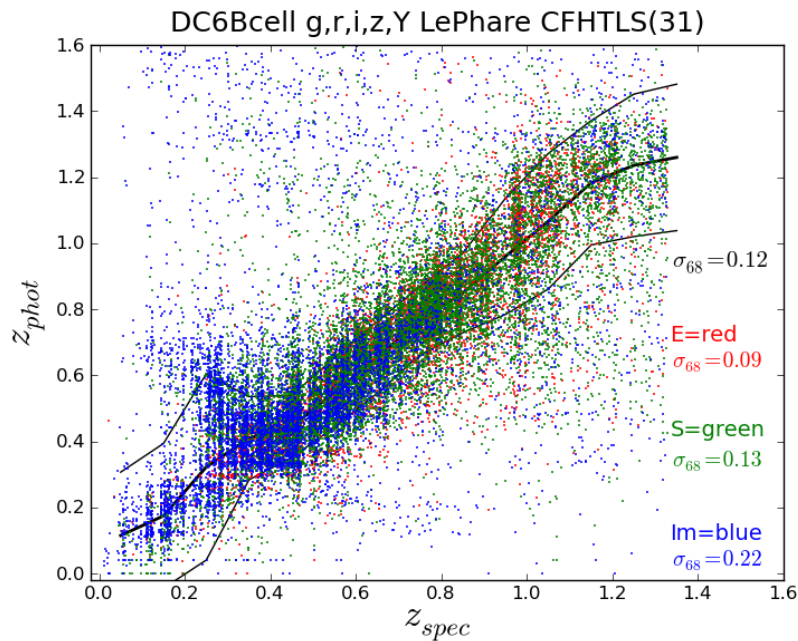
Os resultados provenientes do uso de 21 SEDs do CFHTLS mostram a vantagem de utilização de um conjunto mais amplo cobrindo com melhor resolução as variações espectrais das galáxias. Assim a dispersão  $\sigma_{68}$  é um

pouco reduzida, ficando no limite do requisito do projeto. Além disso, o resultado para o conjunto de SEDs CFHTLS apresenta uma distribuição dos *redshifts* fotométricos mais uniforme.

Nas Figuras 10 e 11 são mostradas as comparações entre os photo-zs obtidos pelo código LePhare e os da simulação DC6B para os diferentes tipos de galáxias atribuídos pelo código. Novamente as Elípticas são mostradas em vermelho, as Espirais (Sbc e Scd) em verde e as Irregulares em azul. Para as SEDs CWW-Kinney, é possível constatar que a maior dispersão é causada por um viés das Elípticas (sistematicamente abaixo da relação 1:1) e pela alta dispersão intrínseca das Irregulares. Já para as SEDs CFHTLS, todos os tipos apresentam baixo viés e o principal tipo causador de dispersão é o das Irregulares, tanto pela presença do desvio localizado em *redshift* espectroscópico  $z_{spec} = [0.2, 0.3]$  quanto pelos chamados erros catastróficos ( $f > 2\sigma$ ).

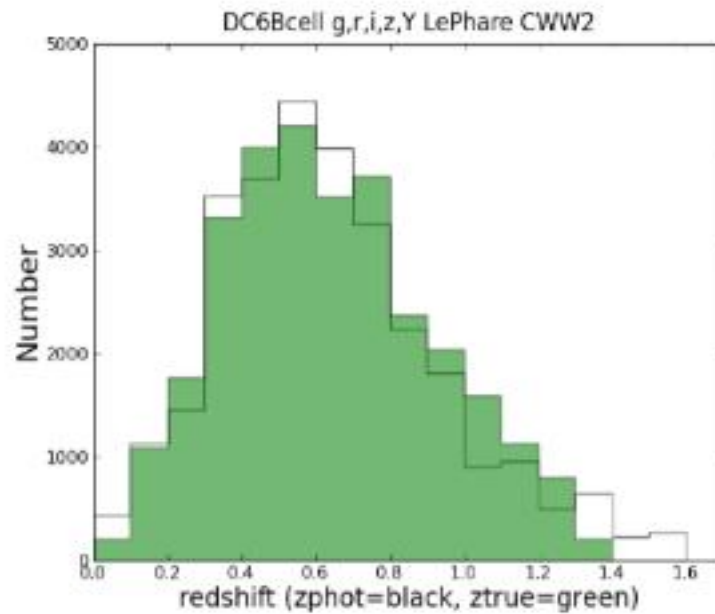


**Figura 10:** *Redshifts* fotométricos ( $z_{phot}$ , LePhare) versus *redshifts* simulados ( $z_{spec}$ , DC6B). Galáxias Elípticas representadas em vermelho, Sbc e Scd em verde, Irregulares Im em azul. As SEDs utilizadas foram as CWW.

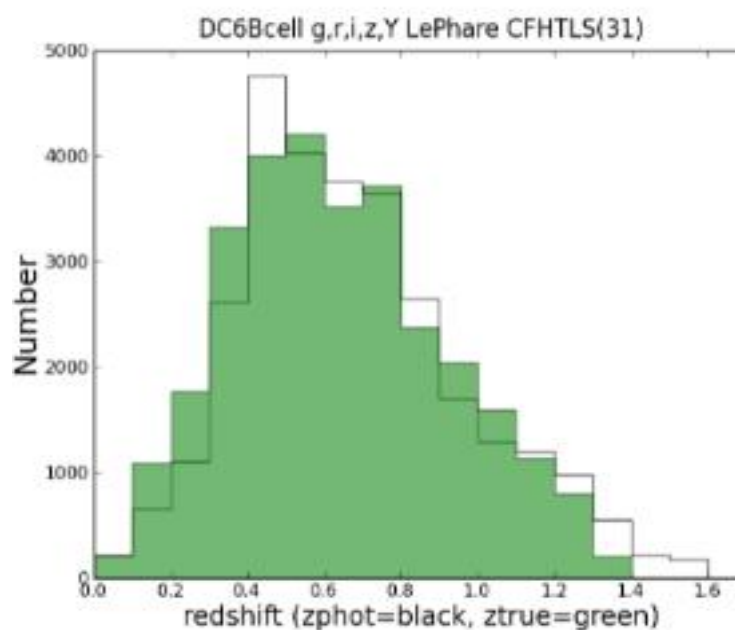


**Figura 11:** *Redshifts* fotométricos ( $z_{phot}$ , LePhare) versus *redshifts* simulados ( $z_{spec}$ , DC6B). Galáxias Elípticas representadas em vermelho, Sbc e Scd em verde, Irregulares Im em azul. As SEDs utilizadas foram as CFHTLS.

Os histogramas das distribuições de *redshifts* para a simulação DC6B são mostrados nas Figuras 12 e 13 para as 6 SEDs CWW-Kinney e para as 21 SEDs CFHTLS, respectivamente. Em ambos os casos a distribuição de photo-zs obtida é bem semelhante à da simulação, com valores de  $\text{rms}(\Delta N/N) \cong 0.01$ , similares aos obtidos com a amostra VVDS.



**Figura 12:** Histograma da distribuição de *redshifts*. *Redshifts* espectroscópicos mostrados em verde, enquanto os fotométricos, determinados pelo LePhare, em preto. SEDs CWW utilizadas.



**Figura 13:** Histograma da distribuição de *redshifts*. *Redshifts* espectroscópicos mostrados em verde, enquanto os fotométricos estão em preto. SEDs CFHTLS utilizadas.

### 3.1.1.3 Amostra VVDS sem a banda u

Uma forma alternativa de avaliar resultados do DES, principalmente no que concerne ao efeito da ausência de magnitudes na banda u, pode ser obtida utilizando o mapeamento VVDS, dele removendo os dados da banda u.

Os resultados de photo-z provenientes da aplicação do código LePhare sobre o conjunto assim obtido é mostrado em comparação com os *redshifts* espectroscópicos na Figura 14. O mesmo conjunto de 21 SEDs CFHTLS, valores de extinção  $E(B-V)$  e *templates* com possibilidade de extinção usados nos resultados da Figura 6 foram utilizados e os resultados são mostrados na Figura 14. Os principais resultados para a amostra VVDS sem a banda u são:

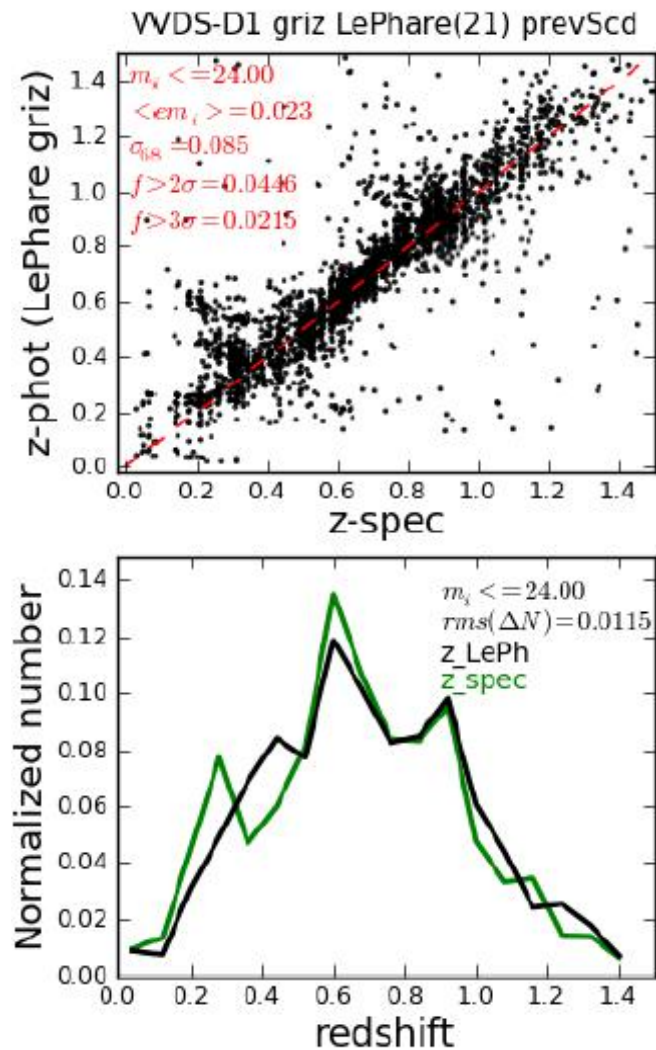
a) o aumento da dispersão para  $\sigma_{68} = 0.085$ , representando um valor 40% superior ao obtido com a banda u, ainda assim inferior ao limite estabelecido pelo projeto;

b) cerca de 4.5% dos objetos possuem erros catastróficos, acima de  $2\sigma$ , representando um valor 30% superior ao obtido com a banda u, ainda assim inferior ao limite requerido pelo projeto;

c) cerca de 2.1% dos objetos tem erros catastróficos superiores a  $3\sigma$ , essencialmente a mesma fração obtida com a banda u, um pouco superior ao limite estabelecido pelo projeto;

d) os histogramas representando as distribuições de photo-zs reproduzem os *redshifts* espectroscópicos com qualidade praticamente idêntica a daqueles obtidos com a banda u;

e) devido ao desvio localizado em *redshift* espectroscópico [0.2,0.3], que artificialmente atribui aos objetos localizados neste domínio valores entre 0.4 e 0.6, a distribuição de photo-zs somente reproduz aquela de *redshifts* espectroscópicos para valores de photo-zs superiores a 0.6.



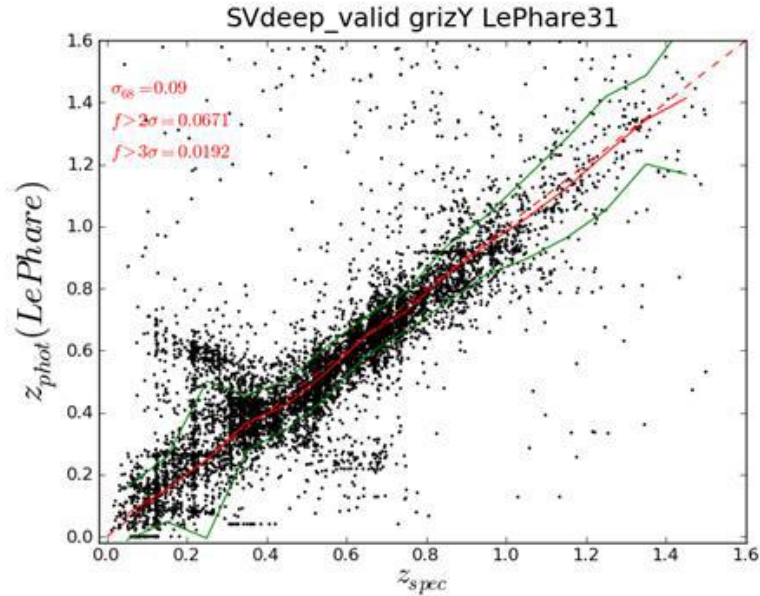
**Figura 14:** Comparação de resultados de photo-zs obtidos com o código LePhare e *redshifts* espectroscópicos da amostra VVDS sem a banda u (painel superior). Comparação das distribuições destes dois tipos de *redshifts* (painel inferior).

### 3.1.1.4 Amostra de Verificação Científica do DES

Também aplicamos o LePhare aos dados disponíveis das primeiras observações do Dark Energy Survey. Estes dados foram utilizados para as primeiras verificações científicas do projeto. O grupo de trabalho em *redshifts* fotométricos do DES elaborou amostras de galáxias com fotometria de boa qualidade e com determinações de *redshifts* espectroscópicos e solicitou aos pesquisadores do projeto que realizassem determinações de *redshifts* fotométricos com seus distintos códigos para posterior comparação. Uma destas amostras atingindo maior profundidade (SVdeep, mencionada no início da seção 3), com cerca de 15500 galáxias, foi utilizada neste trabalho. Como mencionado anteriormente, as magnitudes determinadas no projeto foram obtidas nas bandas g, r, i, z, Y.

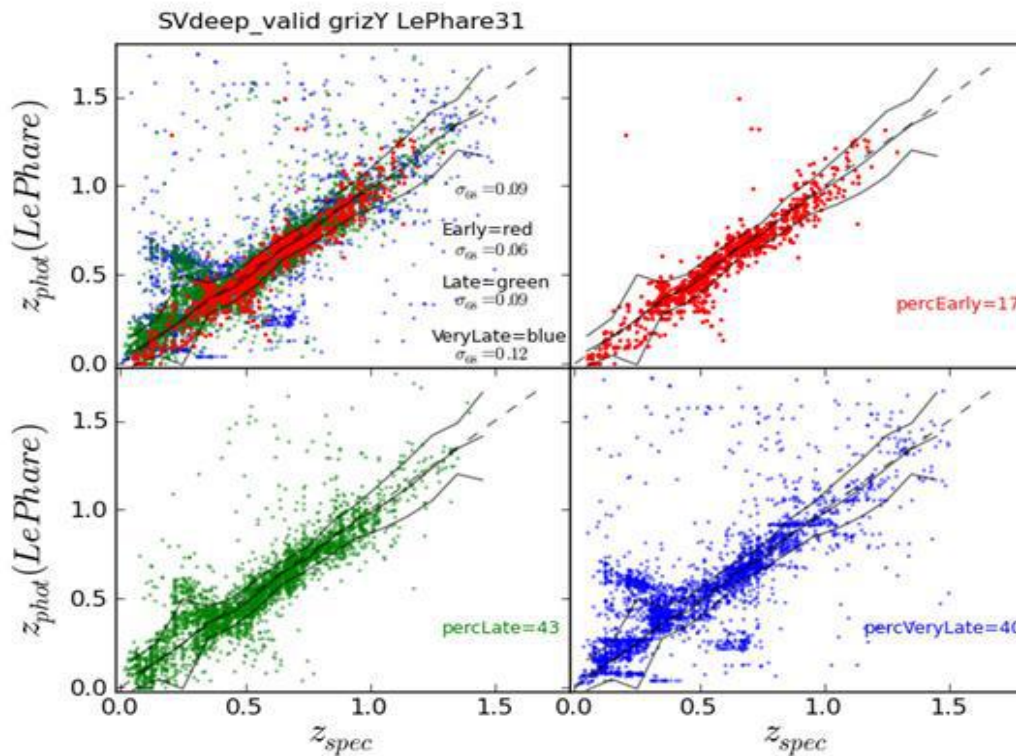
Na Figura 15 é mostrada a comparação dos resultados obtidos com o código LePhare para a amostra SVdeep usando o conjunto reduzido de 21 SEDs CFHTLS e extinção máxima de  $E(B-V) = 0.25$  para tipos Scd e mais *late*. Como pode ser visto, os resultados são bastante semelhantes aos obtidos com a amostra VVDS sem a banda u, tanto qualitativamente quanto nos valores das métricas encontradas:  $\sigma_{68} = 0.09$ ,  $f > 2\sigma = 0.067$  e  $f > 3\sigma = 0.019$ . As metas do DES são atingidas para  $\sigma_{68}$  e  $f > 2\sigma$ .





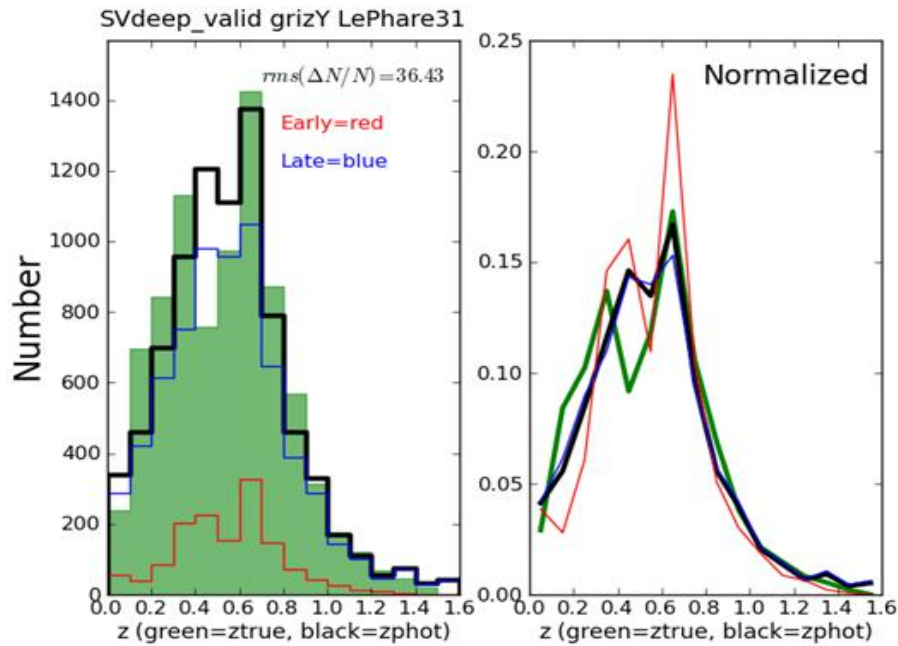
**Figura 15:** Comparação entre *redshifts* fotométricos (LePhare) e *redshifts* espectroscópicas. A linha vermelha tracejada é a relação 1:1, a linha vermelha contínua representa o valor médio da distribuição e as linhas verdes o espalhamento  $\sigma_{68}$  ao longo da distribuição.

Para examinar a qualidade dos *redshifts* fotométricos em função dos diferentes tipos de galáxias, mostramos na Figura 16 as comparações de photo-zs e *redshifts* espectroscópicas para cada tipo e os respectivos espalhamentos. Como pode ser visto, o código LePhare determina *redshifts* fotométricos para os dados do DES com desvio padrão  $\sigma_{68}$  dentro das metas do projeto para todos os tipos. A figura mostra de modo claro que as maiores discrepâncias ocorrem para os tipos Irregulares e galáxias *starbursts*.



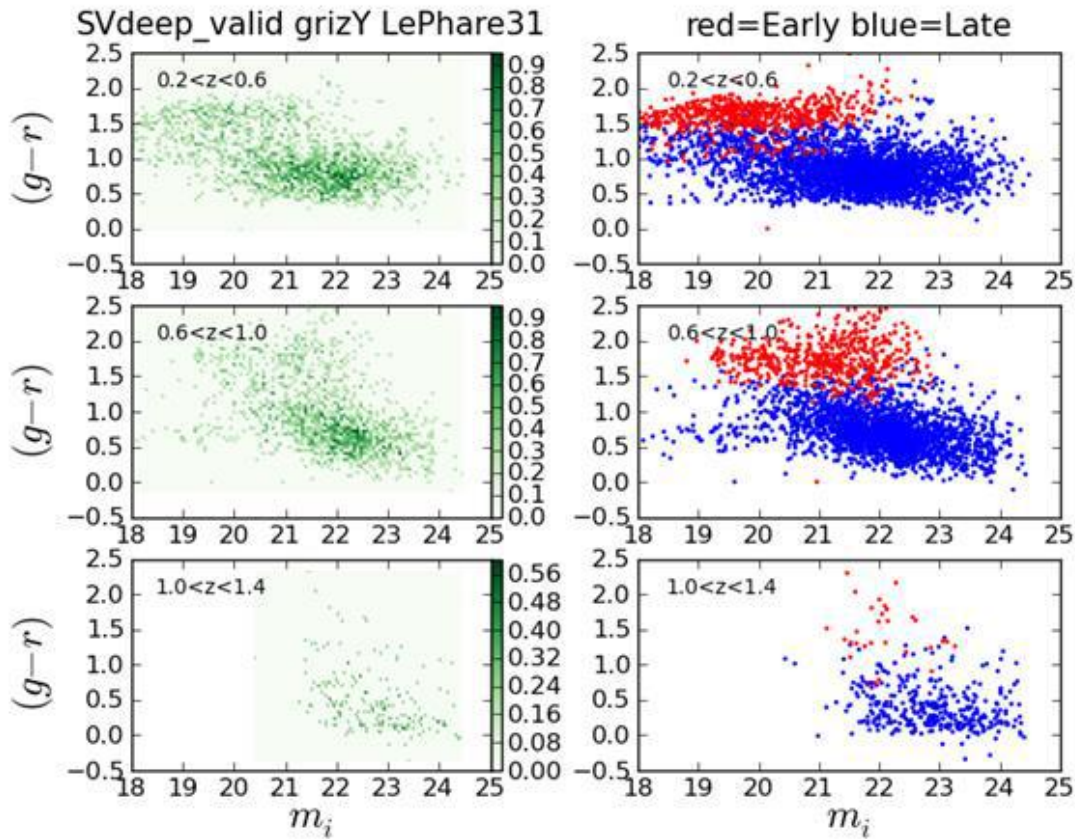
**Figura 16:** Comparação entre *redshifts* fotométricos (obtidos com o LePhare) e *redshifts* espectroscópicos. O painel superior esquerdo mostra toda a distribuição, o superior direito mostra as galáxias Elípticas, o inferior esquerdo mostra as Espirais e o inferior direito mostra as Irregulares e *starbursts*. Cada painel mostra também a percentagem de cada tipo.

A Figura 17 mostra as distribuições de *redshifts* fotométricos (obtidos com o LePhare) e espectroscópicos para os dados do SVdeep. Nela mostramos também as distribuições de diferentes tipos de galáxias: *Early* (Elípticas) e *Late* (Espirais, Irregulares e galáxias *starbursts*). A figura compara esses photo-zs com a distribuição dos *redshifts* espectroscópicos, mostrando os números de objetos e as distribuições normalizadas. Pode-se verificar, como no exame do VVDS sem a banda u, que um acordo aceitável só existe para *redshifts* maiores que 0.6.



**Figura 17:** Distribuição de *redshifts*. Em verde (histograma e linha cheia) é mostrada a distribuição para os *redshifts* espectroscópicos, enquanto em preto (histograma e linha cheia) é mostrada a distribuição de photo-zs. As linhas finas mostram as distribuições de photo-zs para as Elípticas em vermelho e Espirais, Irregulares e starbursts em azul.

A atribuição de tipos de galáxias realizada pelo código LePhare, se feita corretamente, deve apresentar a conhecida bimodalidade (ex. Baldry et al. 2004) entre tipos *Early* e *Late* em diagramas cor-magnitude. Isso foi verificado elaborando-se diagramas da cor ( $g-r$ ) em função da magnitude na banda  $i$ , como mostrado na Figura 18. Nela pode-se verificar a presença da bimodalidade com a identificação clara da chamada sequência vermelha formada pelas galáxias tipo *Early* (Elípticas) para valores de ( $g-r$ ) superiores a 1.5.



**Figura 18:** Diagrama cor  $(g-r)$  versus magnitude na banda  $i$  em diferentes intervalos de *redshift*. Nos painéis da esquerda são mostradas todas as galáxias num mapa de densidade, enquanto nos da direita são mostradas as galáxias separadas nos tipos *Early* e *Late* (de acordo com a classificação do LePhare), apresentadas respectivamente em vermelho e azul.

### 3.1.2 Magnitudes Absolutas determinadas com o LePhare

Uma das principais vantagens de códigos de ajustes de *templates* é sua capacidade de produzir estimativas de correções-K a partir de suas bibliotecas obtidas pelo deslocamento em *redshift* de SEDs de diferentes tipos de galáxias. Assim, além da determinação de tipos, o código fornece estimativas de magnitudes absolutas que constituem parâmetros essenciais para uma grande variedade de estudos de formação e evolução de galáxias.

As magnitudes absolutas  $M$  (numa determinada banda) são calculadas pelo código da forma usual:

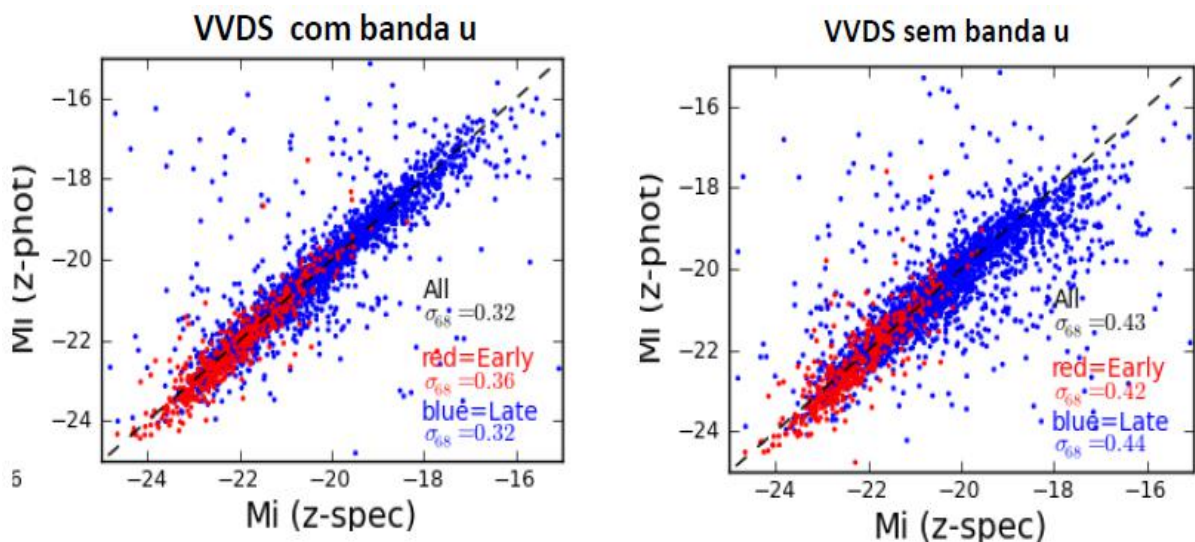
$$M = 25 + m - 5 \log D_L - K,$$

em que  $m$  é a magnitude aparente,  $K$  é a correção- $K$  e a distância de luminosidade (em Mpc) é  $D_L = (1+z) D_c$ , sendo  $D_c$  a distância comóvel calculada através do modelo cosmológico aqui adotado. Os parâmetros  $z$  e  $K$  são definidos pelo ajuste do código LePhare, e  $D_c$  é calculado para a cosmologia adotada.

Como *redshifts* fotométricos apresentam erros consideravelmente superiores aos espectroscópicos (ver figura 2 de Benitez et al. (2009) para uma visualização disto), é importante determinar uma estimativa das incertezas envolvidas nas determinações de magnitudes absolutas. Tendo em vista que a ausência da banda  $u$  nos dados do DES representa um aumento na incerteza dos photo-zs obtidos ( $z < 0.6$ ), é igualmente importante tentar quantificar o quanto essa incerteza afeta as estimativas de magnitudes absolutas determinadas pelo código.

Tomamos inicialmente como valor (pseudo) real da magnitude absoluta de uma galáxia aquele calculado fixando o *redshift* espectroscópico no LePhare. Obviamente este não pode ser considerado um valor real, já que o código escolherá um tipo e uma correção- $K$ , num processo que tem seus erros intrínsecos. Uma comparação com esses valores representará um limite inferior para a incerteza nas determinações de magnitudes absolutas.

Utilizamos os resultados das rodadas de LePhare nas amostras VVDS com e sem a banda u, comparando as magnitudes absolutas de cada uma com os valores pseudo-reais obtidos com os *redshifts* espectroscópicos. Essas comparações estão mostradas na Figura 19, onde os tipos *Early* (Elípticas) e *Late* (Espirais e Irregulares) são discriminados com cores vermelha e azul, respectivamente. O limite inferior para a estimativa de incertezas na magnitude absoluta é de 0.32 mag para a amostra incluindo a banda u, e cresce para 0.43 para a amostra sem a banda u. As incertezas relativas aos tipos não diferem muito desses valores.



**Figura 19:** Comparação de Magnitudes Absolutas estimadas com photo-zs provenientes do código LePhare com as obtidas fixando-se no código o *redshift* espectroscópico, para a amostra VVDS com e sem a banda u.

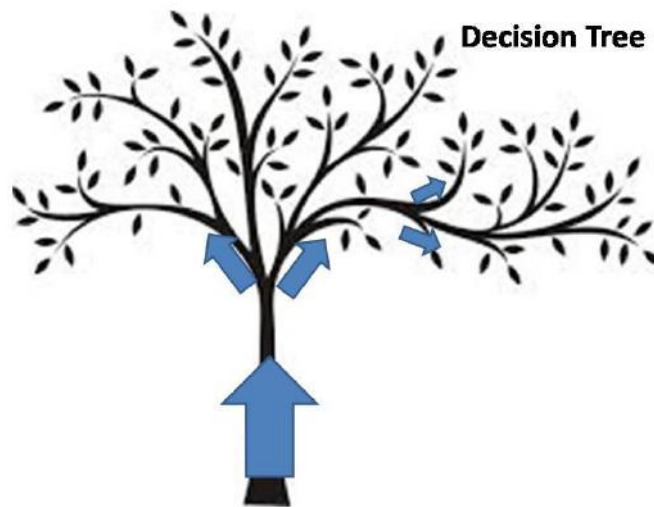
### 3.2 TPZ

O código TPZ, que utiliza árvores de decisão para a determinação de *redshifts* fotométricos, é público, está escrito em linguagem Python, pode ser facilmente obtido na web e disponibiliza recursos para visualização dos resultados. Sua utilização minimamente eficiente requer processamento em paralelo, viabilizado pela rotina pública *mpirun* (Message Passing Interface) escrita pelo MPI Forum. O TPZ faz parte de um pacote mais amplo (MLZ) de “máquinas de aprendizagem”, incluindo outro código denominado SOMz. O TPZ utiliza uma amostra de treinamento com dados fotométricos e *redshifts* espectroscópicos, e com ela serão construídas as árvores de decisão, a partir das quais serão determinados os *redshifts* fotométricos.

Essencialmente, o TPZ opera como esquematizado na Figura 20. Toda a amostra de treinamento está inicialmente contida no tronco de uma árvore e a cada bifurcação para caules mais finos é aplicado um critério estatístico subdividindo a amostra em 2 partes, de modo que elas sejam diferentes ao máximo. Assim, numa folha (final do processo de subdivisão) estarão objetos cujos parâmetros (magnitude, cores, etc) e *redshift* espectroscópico estão estatisticamente associados. Amostras fotométricas sem *redshifts* submetidas ao código têm suas magnitudes e cores estatisticamente classificadas pela árvore assim definida e finalizam com associações de suas magnitudes e cores a diferentes *redshifts*. Com essas distribuições de *redshifts* (pdf) o TPZ calcula, para cada objeto, photo-zs como:

photo-z *mean* = média da sua pdf

photo-z *mode* = valor do pico da sua pdf



**Figura 20:** Esquema ilustrando o funcionamento do TPZ. Amostra total inicial no caule da árvore, sendo subdividida nos entroncamentos de caules de forma a maximizar as diferenças.

O processamento do código TPZ é definido num arquivo de configuração incluindo:

- a) Definição das amostras de treinamento e daquela a ter photo-zs determinados;
- b) Definição das colunas contendo parâmetros de entrada (magnitude, cores, etc...);

O arquivo permite uma grande variedade de escolhas de parâmetros de execução e optamos por variar apenas as seguintes:

- c) Escolha dos limites de *redshift* e número de *bins*;



d) Escolha do número de árvores, do número de florestas (realizações randômicas dessas árvores) e do número final mínimo de objetos nas folhas.

Baseados em resultados previamente obtidos pelo grupo brasileiro no DES, verifica-se que os photo-zs reproduzem com boa estatística os *redshifts* espectroscópicos quando:

Parâmetros de entrada: 4 magnitudes (g,r,i,z) e 3 cores (g-r), (r-i) e (i-z)

Domínio de *redshift*: 0 a 3 em intervalos  $dz = 0.01 - 0.02$

Número de árvores: 50 a 100

Número de florestas: 1 a 10

Número mínimo de objetos nas folhas: 3 a 6

O número total de árvores (número de árvores  $\times$  número de florestas) mínimo deve ser igual ou superior a 150 para garantir boas determinações de photo-zs. Para os resultados apresentados a seguir, são mostrados os valores *mean* para o photo-z. Os valores *mode* são usualmente binados já que se referem a um valor obtido no centro do *bin* de *redshift* onde ocorre o máximo da pdf.

Um ponto importante na verificação da eficiência de códigos de treinamento é que a amostra de treinamento e a de validação não devem ter objetos em comum. Se usarmos, por exemplo, a mesma amostra para treinamento e validação, as galáxias da amostra de validação têm grande probabilidade estatística de seguir a mesma subdivisão nos caules das árvores e maior a chance de finalizarem em folhas semelhantes às aquelas que elas mesmo definiram no processo de treinamento. Desta forma, a

comparação entre photo-zs e *redshifts* espectroscópicos apresentará dispersão artificialmente mais baixa que a realidade.

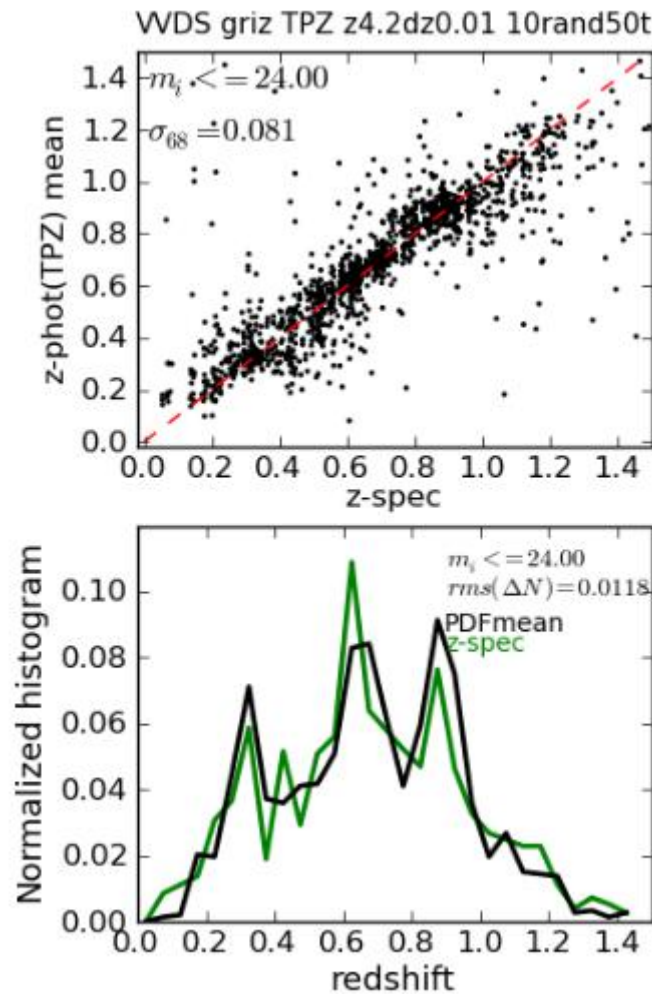
### 3.2.1 Photo-zs determinados com o TPZ

#### 3.2.1.1 Amostra VVDS sem a banda u

Aplicamos o código TPZ na mesma amostra VVDS para a qual as magnitudes na banda u foram descartadas. Como mencionado acima, para evitar resultados artificialmente melhores, subdividimos a amostra de forma aleatória em cerca de 1500 galáxias para treinamento e 1500 para validação. Os resultados comparando os photo-zs obtidos com o código TPZ e os respectivos *redshifts* espectroscópicos, para a amostra VVDS de validação, são mostrados na Figura 21. Como a amostra é pequena usamos um número total de árvores grande para melhorar a estatística: 10 florestas de 50 árvores.

No painel superior desta figura, pode-se constatar que a dispersão ( $\sigma_{68} = 0.081$ ) obtida é ligeiramente inferior àquela obtida com o código LePhare (Figura 14). As frações de erros catastróficos são  $f > 2\sigma = 0.023$  e  $f > 3\sigma = 0.014$ . Em particular deve-se destacar que esse resultado satisfaz a todos os requisitos do DES, incluindo a fração de erros acima de  $3\sigma$ .

Ademais, uma melhora significativa é alcançada para o domínio de  $z\text{-phot} < 0.6$ . Este último aspecto pode ser verificado também no painel inferior da Figura 21 com um melhor acordo da distribuição de *redshifts* (photo-z e espectroscópico) neste domínio.

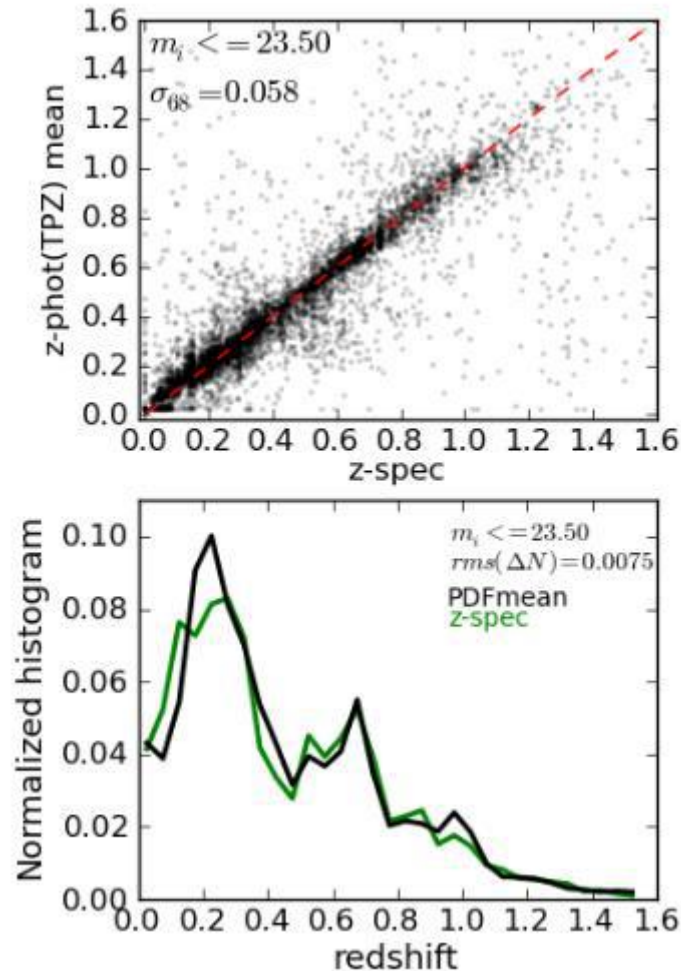


**Figura 21:** Comparação de resultados de photo-zs obtidos com o código TPZ e *redshifts* espectroscópicos da amostra VVDS sem a banda u (painel superior). Comparação das distribuições destes dois tipos de *redshifts* (painel inferior).

### 3.2.1.2 Amostra da Verificação Científica do DES

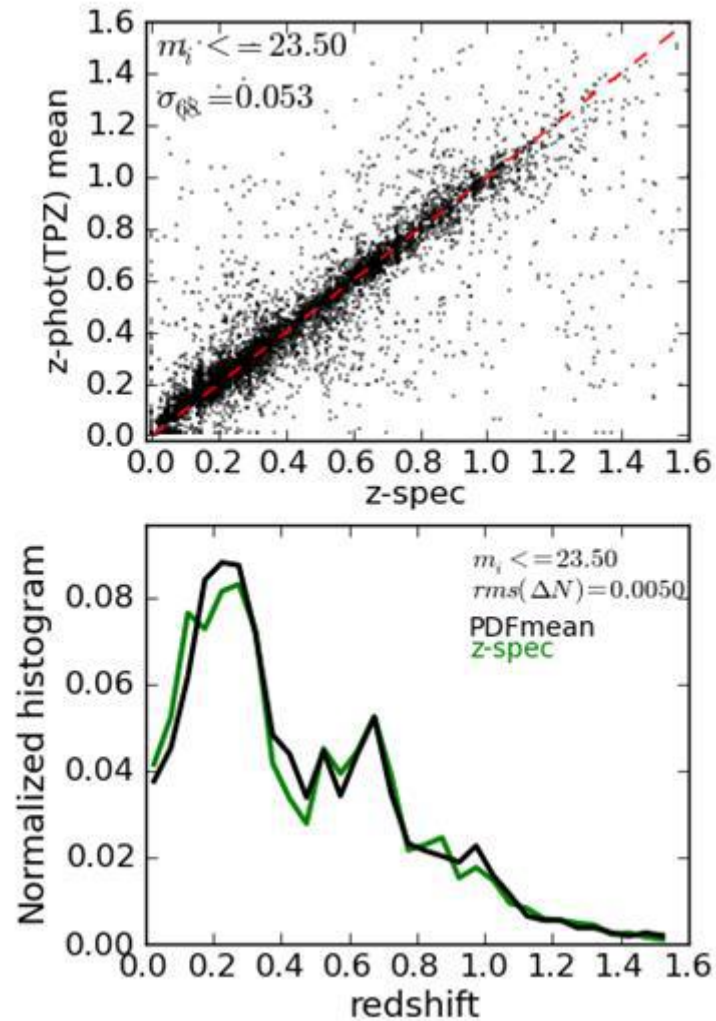
Na Figura 22 é mostrada a comparação dos resultados de photo-zs obtidos com o código TPZ para a amostra SVdeep. Neste caso, tanto a amostra de treinamento quanto a de validação utilizadas foram aquelas definidas pelo Grupo de Trabalho de *Redshifts* Fotométricos da colaboração DES (Sanchez et al. 2014, mencionadas no início da seção 3).

Variamos um pouco as configurações de entrada para realizar alguns poucos testes. Um resultado típico é mostrado na Figura 22 com 4 florestas de 25 árvores.



**Figura 22:** Comparação de resultados de photo-zs obtidos com o código TPZ (4 florestas de 25 árvores) e *redshifts* espectroscópicas da amostra SVdeep (painel superior). Comparação das distribuições destes dois tipos de *redshifts* (painel inferior).

A execução do TPZ com melhor resultado foi feita com 3 florestas de 50 árvores, com o número mínimo de 3 objetos em cada folha, e é mostrada na Figura 23. Os resultados mostram determinações de photo-z melhores do que os apresentados anteriormente, com uma relação bastante linear e dispersão  $\sigma_{68} = 0.053$ . As frações de galáxias com erros catastróficos são  $f > 2\sigma = 0.039$  e  $f > 3\sigma = 0.024$ , ficando esta última acima do limite desejado pelo projeto.



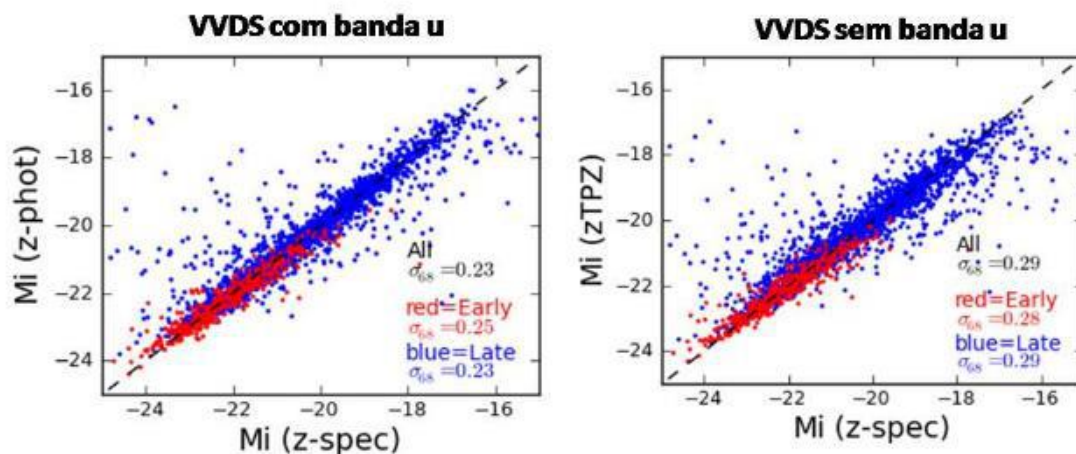
**Figura 23:** Comparação de resultados de photo-zs obtidos com o código TPZ (3 florestas de 50 árvores) e *redshifts* espectroscópicos da amostra SVdeep (painel superior). Comparação das distribuições destes dois tipos de *redshifts* (painel inferior).

Deve-se notar que, além de possuir uma maior amostra de treinamento, a amostra SVdeep possui erros de magnitude menores que a amostra VVDS. Esses fatores devem contribuir para as melhores determinações de photo-z.

### 3.2.2 Magnitudes Absolutas no LePhare com photo-zs do TPZ

Como os resultados de photo-z provenientes do TPZ se mostram superiores aos determinados com o LePhare, especialmente no domínio  $z < 0.6$ , consideramos a alternativa de fixar os valores obtidos com TPZ no LePhare, deixando este último realizar a determinação de tipos, correções-K e magnitudes absolutas.

Isto foi feito tanto na amostra VVDS com a banda u como na amostra VVDS descartando-se as magnitudes na banda u, da mesma forma mostrada na seção 3.1.2 para os resultados puramente determinados com o LePhare. Os resultados das determinações de magnitudes absolutas obtidas com photo-zs do TPZ comparados com aquelas obtidas com os *redshifts* espectroscópicos são mostrados na Figura 24, respectivamente para as amostras com e sem a banda u. A figura mostra esta comparação discriminando os tipos *Early* e *Late*, cujas dispersões são essencialmente as mesmas da amostra total.



**Figura 24:** Comparação de Magnitudes Absolutas estimadas com photo-zs provenientes do código TPZ com as obtidas fixando-se no código o *redshift* espectroscópico, para a amostra VVDS, com e sem a banda u.

Os resultados de photo-zs determinados com o TPZ são menos sensíveis à ausência da banda u, resultando num aumento do espalhamento de apenas 0.06 mag, cerca de metade daquele verificado nos resultados provenientes do LePhare, mostrado na Figura 19.

## 4 Conclusões e comentários finais

Examinamos nesta monografia dois métodos de determinação de *redshifts* fotométricos representativos das duas principais classes destes códigos: o LePhare, que utiliza ajuste de *templates*, e o TPZ, que usa processo de treinamento.

Uma clara diferença de resultados provém da existência ou não de dados na banda u na amostra sendo analisada. A ausência dela reduz a eficiência dos códigos pela incerteza causada em baixos *redshifts* ( $z < 0.6$ ). No caso do LePhare esse efeito é particularmente forte, pois utilizando ajuste de *templates*, as variações das SEDs no ultravioleta (como o aumento de fluxo causado por formação estelar) em diferentes *redshifts* no domínio  $z < 0.6$  não são bem determinadas devido à falta da banda u, e então não podem ser usadas para os ajustes que definem o photo-z. Com a perda dessa definição, *templates* de tipos *Late*, com diferentes possibilidades de extinção, podem ajustar melhor as magnitudes de uma galáxia tipo *Early*. Esse aspecto é o causador das discrepâncias encontradas localmente no domínio  $z < 0.6$ . Tipicamente, a dispersão  $\sigma_{88}$  da comparação photo-z com *redshifts* espectroscópicos cresce de 0.06, para uma amostra (de 5 bandas óticas) incluindo a banda u, para 0.085, sem a banda u.

Por outro lado, é importante lembrar que o projeto DES teve sua estratégia observacional definida para atingir objetivos específicos no estudo da energia escura. Com isso, seu alvo principal é o Universo acima de  $z = 0.6$ , e por isso os filtros do projeto foram escolhidos como grizY.

Sendo um código de ajuste de *templates*, o LePhare tem sua eficiência dependente do conjunto de SEDs escolhido. A utilização de um conjunto pequeno de *templates*, como CWW-Kinney (6 SEDs), pode não



cobrir toda a diversidade de cores de amostras reais, além de fornecer valores de photo-zs com baixa resolução. O emprego de um conjunto maior, como aquele das análises do CFHT *Legacy Survey*, mesmo aqui reduzido para 21 SEDs, mostra a importância de se utilizar uma maior variedade de *templates*.

A eficiência de determinação de photo-zs é maior para as galáxias tipo *Early* em comparação com as tipo *Late*, um resultado conhecido e natural, já que o espectro das primeiras é mais regular e marcado pela quebra em 4000 Angstroms. Entretanto, é provável que algumas Elípticas sejam erradamente classificadas como um tipo *Late* com extinção, num *redshift* incorreto, devido à ausência da banda u, conforme mencionado acima. Além disso, contribuem para a maior incerteza no photo-z das galáxias tipo *Late* as diferentes taxas de formação estelar, que implicam em diferenças nas magnitudes, causadas pela linhas de emissão resultantes desse processo.

O TPZ resulta em determinações de photo-zs de melhor qualidade do que o LePhare, com base nas estatísticas mostradas. Um ajuste fino na configuração de execução do código, como o número de florestas e árvores, permite a obtenção de photo-zs que guardam linearidade com os *redshifts* espectroscópicos e uma significativa redução do problema causado pela ausência da banda u. Desta forma foi possível obter uma dispersão  $\sigma_{68} = 0.053$  para os dados de verificação científica do DES.

Como o código LePhare representa uma ferramenta de grande valor para a determinação de parâmetros físicos das galáxias, uma alternativa natural foi a ingestão dos resultados de photo-zs determinados com TPZ na execução do LePhare, mostrada na seção 3.2.2.

Dentro do contexto desse procedimento, procuramos estimar limites inferiores para as incertezas nas determinações de magnitudes absolutas com o código LePhare, e alguns aspectos devem ser ressaltados:

- a) O primeiro é que essas incertezas mínimas são da ordem de 0.25 a 0.45 mag, dependendo de se a amostra tem ou não a banda u e de se os photo-zs são determinados pelo LePhare ou pelo TPZ;
- b) O segundo é que a estimativa do aumento desses erros em função da ausência da banda u nos dados do DES é de 0.06 mag (photo-zs produzidos pelo TPZ) ou 0.1 mag (photo-zs produzidos pelo LePhare);
- c) A estimativa de um valor pseudo-real para as magnitudes absolutas usando dados do DES seria muito imprecisa devido à ausência da banda u. Para realizar essas estimativas utilizamos a amostra VVDS, que embora pequena, é semelhante em filtros ao DES, porém contém a banda u, que foi mantida ou retirada para realizar as comparações desejadas.

Com todas essas restrições, as estimativas devem ser consideradas grosseiras e como limites inferiores. A literatura tem apontado que, por exemplo, correções-K representam uma das principais fontes de erros nas determinações de magnitudes absolutas. Por outro lado, a colaboração DES iniciou um projeto observacional para obtenção de dados na banda u, o que significará um considerável aprimoramento nas determinações de photo-zs.

Em todas as estimativas de erros, a ingestão de photo-zs (de melhor qualidade) do TPZ no LePhare produziu os menores erros para as magnitudes absolutas. O conjunto de resultados aqui apresentado indica

que este é um bom procedimento para determinar photo-zs e parâmetros físicos de galáxias.

Para finalizar, devemos mencionar que outros testes foram realizados durante o período de execução deste trabalho, mas não foram aqui apresentados para manter uma coerência temática e resultados concretos. Entre eles estão:

a) Examinamos também os resultados provenientes do código de redes neurais ANNz, e para amostras sem a banda u, ele apresenta grande concentração de photo-zs no entorno de 0.4, além de desvios sistemáticos nos limites inferior e superior do domínio de *redshifts*.

b) Executamos o LePhare utilizando SEDs sintéticas como o PEGASE-2 e os modelos de Maraston (2005). Os problemas causados pela ausência da banda u em geral se intensificam com as SEDs sintéticas.

c) Iniciamos uma avaliação da cobertura do espaço de cores (g-r) x (r-i) por alguns conjuntos de SEDs e constatamos que as SEDs CFHTLS não cobrem todo o espaço de amostras reais. Por outro lado, resultados preliminares com as SEDs utilizadas no mapeamento COSMOS se mostraram mais eficazes em reproduzir a diversidade de cores de amostras reais. A determinação de um melhor conjunto de SEDs para utilização no código LePhare é uma iniciativa valiosa para aprimorar essa ferramenta tão útil no estudo de evolução de galáxias.

## 5 Bibliografia

Albrecht A. et al., 2006, arXiv:astro-ph/0609591

Arnouts S. et al., 2002, MNRAS, 329, 355

Baldry I. et al., 2004, ApJ, 600, 681

Bender R., Appenzeller I., Bohm A., al E., 2001, Deep Fields, ed. S. Cris. A. Renzini, R. E. Williams, 96

Benitez N., 2000, ApJ, 536, 571

Benitez N. et al., 2009, ApJ, 691, 241

Bertin E., Arnouts S., 1996, A&A, 117, 393

Bolzonella M., Miralles J., Pelló R., 2000, A&A, 363 476

Bonnett C., 2013, ArXiv e-prints

Bordoloi R., Lilly S. J., Amara A., 2010, MNRAS, no

Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503

Brimioulle F., Seitz S., Lerchster M., Bender R., Snigula J., 2013, MNRAS, 432, 1046

Brodwin M. et al., 2006, ApJ, 651, 791

Bruzual A. G., Charlot S., 1993, ApJ, 405, 538

Calzetti, D. et al., 2000, ApJ, 533, 682

Carrasco Kind M., Brunner R. J., 2013, MNRAS, 432, 1483

Carrasco Kind M., Brunner R. J., 2014, arXiv:1403.0044v1

Coe D., Bentez N., Sanchez S. F., Jee M., Bouwens R., Ford H., 2006, AJ, 132, 926

Coleman G. D., Wu C.-C., Weedman D. W., 1980, ApJS, 43, 393

Colless M. et al., 2001, MNRAS, 328, 1039

Collister A. A., Lahav O., 2004, PASP, 116, 345

da Costa, L. N. et al., 1988, ApJ, 327, 544

De Vicente, J. 2014 (comunicação privada)

Davis, M. Huchra J., 1982, ApJ, 254,437

Dawson K. S. et al., 2013, AJ, 145, 10

Drinkwater M. J. et al., 2010, MNRAS, 401, 1429

Feldmann R. et al., 2006, MNRAS, 372, 565

Fioc M., Rocca-Volmerange B., 2000, arXiv: astro-ph/9912179

Flaugher B. et al., 2005, AJ, 150, 150

Gerdes D. W., Sypniewski A. J., McKay T. A., Hao J., Weis M. R., Wechsler R. H., Busha M. T., 2010, *ApJ*, 715, 823

Greisel N., Seitz S., Drory N., Bender R., Saglia R. P., Snigula J., 2013, *ApJ*, 768, 117

Ilbert O. et al., 2006, *A&A*, 457, 841

Kaiser N., Tonry J. L., Luppino G. A., 2000, *PASP*, 112, 768

Kinney A. L., Calzetti D., Bohlin R. C., McQuade K., Storchi-Bergmann T., Schmitt H. R., 1996, *ApJ*, 467, 38

Le Fèvre O. et al., 2004, *A&A*, 428, 1043

Mannucci F., Basile F., Poggianti B., Cimatti A., Daddi E., Pozzetti L., Vanzì L., 2001, *MNRAS*, 326, 745

Maraston C., 2005, *MNRAS*, 362, 799

Oyaizu H. et al., 2008, *ApJ*, 674, 768

Perlmutter S. et al., 1990, *ApJ*, 571, 565

Prevot M. L. et al., 1984, *A&A*, 132, 389

Pyle, 1999, *Data Preparation for Data Mining*. pp. 271-274, 355-359

Riess A. et al., 1990, *AJ*, 116, 1009

Sanchez C. et al., 2014, MNRAS, 445, 1482

Scoville N. et al., 2007, ApJS, 172, 1

Sypniewski A., 2014, PhD thesis, University of Michigan

Tipping, 2001, JMLR, 1, 211

Tyson J. A. et al., 2003, NuPhS, 124, 21

York D. G. et al., 2000, AJ, 120, 1579