

Método bayesiano de casamento espectral de galáxias

William Schoenell

Orientador:

Prof. Dr. Roberto Cid Fernandes Jr.

Coorientadora:

Dra. Natalia Vale Asari

• • •

Universidade Federal de Santa Catarina
Centro de Ciências Físicas e Matemáticas
Programa de Pós-graduação em Física

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Física da UFSC em preenchimento parcial dos requisitos para obtenção do título de Mestre em Física.

Trabalho financiado pelo Instituto Nacional de Ciência e Tecnologia de Astrofísica e CAPES.

Resumo

Em espectros de galáxias, encontramos duas características principais: linhas de emissão, que nos dizem sobre o gás, e o contínuo mais as linhas de absorção, que nos dizem sobre o conteúdo estelar. Nos últimos anos, desenvolvemos um código de síntese espectral, o STARLIGHT, o qual utilizamos para realizar a síntese de 926246 galáxias do SDSS com modelos de população estelar modernos, nos dando importantes propriedades físicas (massas, históricos de formação estelar, etc.) para estudar como galáxias se formam e como elas evoluem no tempo. Linhas de emissão também são estudadas através da sua medida no espectro residual. Por outro lado, a fotometria é muito mais limitada já que não nos permite extrair todas estas propriedades que a resolução da espectroscopia nos traz. Contudo, com os recentes avanços dos *redshift surveys* com filtros muito estreitos ($\approx 100 \text{ \AA}$, como ALHAMBRA, JPAS e DES), nos interessaria saber com qual precisão podemos extrair informação das propriedades físicas e do gás de galáxias a partir de dados fotométricos.

Nosso objetivo neste trabalho é responder, usando dados fotométricos simulados, a seguinte questão: O que podemos aprender sobre galáxias a partir desta nova geração de *surveys*? Por exemplo, podemos medir idade e metalicidade estelares? Podemos separar galáxias de formação estelar daquelas que possuem um núcleo ativo? Podemos medir linhas de emissão, abundâncias nebulares e extinção? Com qual precisão?

Para isto, selecionamos uma amostra de cerca de 300 mil galáxias do SDSS e a dividimos em dois grupos: uma amostra 200 mil objetos e uma base de 100 mil. Corrigimos o espectro para o referencial de repouso ($z = 0$) e convertemos para os filtros do JPAS. Utilizando uma abordagem estatística, calculamos uma função distribuição de probabilidades (PDF) para cada propriedade para cada objeto em relação à base. Uma vez que temos todas as propriedades de todas as galáxias do banco de dados do STARLIGHT-SDSS, podemos compará-las com os resultados obtidos dos estimadores da PDF (média,

mediana, etc).

Nossos resultados comprovam que podemos recuperar a idade média das galáxias com uma boa margem de erro ($\sigma \approx 0.1 - 0.2$ dex) e, de forma similar, também outras propriedades físicas, como razão massa-luminosidade, metalicidade, etc. Além disso, nosso **principal resultado** é que **podemos recuperar intensidades e razões de linhas de emissão** com precisão similar. Isto torna nosso método inédito em comparação aos outros métodos no mercado que analisam fotometria, e mostra que, do ponto de vista de estudos de galáxias, os novos *surveys* serão muito mais úteis que o esperado.

Abstract

On galaxy spectra, one can find mainly two features: emission lines, which tell us about the gas content, and the continuum plus absorption lines, which tell us about the stellar content. In the past years, we developed a spectral synthesis code called STARLIGHT which we used to fit the stellar populations of 926246 galaxies from the SDSS with modern evolution population synthesis models, giving us important physical properties (masses, star formation histories, etc.) to study how galaxies form and evolve in time. Emission lines are also studied, after measuring them from the starlight subtracted spectrum. Photometry, on the other hand, is much more limited, as it does not allow us to confidently retrieve all those physical properties that the resolution of spectroscopy brings us. However, with the recent development of redshift surveys with ultra-narrow filters ($\approx 100 \text{ \AA}$, like ALHAMBRA, JPAS and DES), it is interesting to know how confidently we can retrieve information of the physical properties and the gas content of galaxies from photometric data.

Our aim in this work is to answer, using simulated photometry data, the following general question: What can we learn about galaxies from these new generation surveys? For instance, can we measure stellar age and metallicity distributions? Can we separate star-forming galaxies from AGN? Can we measure emission lines, nebular abundances and extinction? With what precision?

To accomplish this, we selected a sample of about 300k galaxies from the SDSS and divided them in two groups: 200k Objects and a Base of 100k. We corrected the spectra to $z = 0$ and converted them to filter fluxes. Using a statistical approach, we calculated a Probability Distribution Function for each property of each Object and the Base. Since we have the properties of all the data from the STARLIGHT-SDSS database, we could compare them with the results obtained from summaries of the PDF (mean, median, etc).

Our results shows that we retrieve the weighted average of the log of the galaxy age with a good error margin ($\sigma \approx 0.1 - 0.2$ dex), and similarly for the physical properties such as mass-to-light ratio, mean stellar metallicity, etc. Furthermore, our **main result** is that **we can derive emission line intensities and ratios** with similar precision. This makes this method unique in comparison to the other methods on the market to analyze photometry data and shows that, from the point of view of galaxy studies, future photometric surveys will be much more useful than anticipated.

Sumário

1	Introdução	1
1.1	O STARLIGHT e o mega-levantamento astronômico SDSS	3
1.2	O mega-levantamento J-PAS e este trabalho	4
2	Do SDSS ao J-PAS	9
2.1	O megalevantamento SDSS	9
2.2	O megalevantamento J-PAS	13
2.3	Conversão de Espectros do SDSS para o J-PAS	16
2.4	Espectros de erros para o espectro J-PAS	16
2.5	O J-espectro	19
3	Método	21
3.1	Introdução	21
3.2	Definições	24
3.3	Casamento Espectral: Figura de mérito e pesos	24
3.4	Casamento Espectral: Exemplos	26
3.5	Formalismo Estatístico	31
4	Amostras, bases e detalhes técnicos	37
4.1	Amostra Geral	37

4.2	Bases	38
4.2.1	Base de Propriedades Físicas: Diagrama “cor-magnitude”	39
4.2.2	Base de Linhas de Emissão: Diagrama WHAN	42
4.3	Otimizando o fator f_L	45
4.4	Estimadores: Resumos da Função Distribuição de Probabilidades	48
5	Resultados	51
5.1	Propriedades de interesse	51
5.2	Resultados: Propriedades Físicas	53
5.3	Resultados: Linhas de Emissão	55
5.4	Resumo	59
6	Conclusões e Perspectivas	61
6.1	Resumo do trabalho e seus principais resultados	61
6.2	Perspectivas	63
	Referências Bibliográficas	I

Lista de Figuras

1.1	Uma modo de analisar os J-Espectros.	6
2.1	Projeção equatorial das regiões cobertas pelo levantamento espectroscópico do SDSS.	11
2.2	Exemplo de um prato de fibras do SDSS.	12
2.3	Desenho conceitual do mosaico de CCDs que equipará o telescópio de 2,52m no OAJ.	14
2.4	Curvas de transmissividade utilizadas neste trabalho.	15
2.5	Exemplos de J-espectros normalizados.	20
3.1	Esquema gráfico do método proposto para abordar o problema da baixa resolução.	23
3.2	Quatro exemplos de J-espectros observados e seus melhores ajustes.	27
3.3	Quatro espectros para os exemplos da figura anterior.	28
3.4	Exemplo de J-espectro observado e seus 5 melhores ajustes em χ^2	29
3.5	Outro exemplo de J-espectro observado e seus 5 melhores ajustes em χ^2	30
3.6	Figura retirada de Gallazzi et al. (2005) mostrando quatro exemplos de PDFs em relação à idade média.	32
3.7	Função distribuição de probabilidade para $\langle \log t_\star \rangle_L$ do objeto da figura 3.4.	35
3.8	Função distribuição de probabilidade para $\log[N_{ii}]/H_\alpha$ do objeto da figura 3.4.	36

4.1	Diagrama cor/magnitude extraído de Baldry et al. (2004).	40
4.2	Diagrama cor/magnitude para a amostra de galáxias selecionadas como base.	41
4.3	Diagrama WHAN e classificação de galáxias.	43
4.4	Diagrama WHAN para a amostra de galáxias selecionadas como base. . .	44
4.5	Dispersão da diferença entre entrada e saída para diferentes fatores f_L e diferentes propriedades para a base “diagrama cor-magnitude”.	46
4.6	Dispersão da diferença entre entrada e saída para diferentes fatores f_L e diferentes propriedades para a base “WHAN”.	47
5.1	Histogramas saída menos entrada para as propriedades físicas de interesse	54
5.2	Histogramas saída menos entrada para as linhas de emissão	57
5.3	Histogramas saída menos entrada para as razões de linhas de emissão . .	58

Lista de Tabelas

5.1	Propriedades analisadas, suas unidades e descrições.	52
5.2	Melhor f_L , média, mediana, desvio padrão para $\Delta p = p_{saida} - p_{entrada}$, assim como mínimo e máximo para as propriedades físicas analisadas com a base CMD	55
5.3	Tabela igual a 5.2, mas para base WHAN	55
5.4	Tabela igual a 5.2, agora para linhas de emissão	56
5.5	Mesma tabela anterior (5.4), mas para base WHAN	59

Capítulo 1

Introdução

A principal motivação para o estudo da astronomia está ligada à curiosidade do homem em saber qual o seu lugar no Universo. De onde veio, para onde vai. Por isto, o estudo da astronomia muito se confunde com o desenvolvimento da civilização e tem papel fundamental nas mais diferentes civilizações antigas. Maias, incas, astecas, egípcios e muitos outros povos tinham a astronomia como centro para o seu desenvolvimento. Os astrônomos nestas sociedades tinham papel fundamental. Esta curiosidade é observada também quando, com grande frequência, assistimos reportagens na televisão sobre as fascinantes descobertas, jornais divulgam encartes com as pesquisas recentes no campo, internautas cada vez mais consultam imagens do Universo nos sites de grandes telescópios baseados na terra e no espaço.

A astronomia moderna, guiada por equipamentos que ampliam nosso poder de visão, os telescópios, surgiu com Galileu Galilei, em 1609. Com a simples ideia de utilizar um equipamento inventado para que militares vejam tropas inimigas a grande distância para olhar o cosmos, Galileu iniciou a mudança mais radical na nossa visão do que é o Universo. Esta mudança foi tão radical e brusca para os padrões que deixou muita gente da época confusa, descontente e Galileu quase pagou com sua vida por suas descobertas.

Durante os séculos seguintes, a astronomia teve um crescimento incrível. Estrelas, cometas, planetas, luas, aglomerados de estrelas e estranhos objetos, chamados na época de nebulosas espirais, foram descobertos. Diferentes astrônomos realizaram observações destas intrigantes nebulosas. Como, por exemplo, Charles Messier que, no século 18, desenvolveu um catálogo com 109 destas¹ e, anos mais tarde, William Herschel com um

¹Messier estava interessado em observação de cometas, não em galáxias. Catalogou elas como objetos que não eram cometas.

catálogo de mais de 5000 destes objetos.

Até o início do século passado, a natureza destas nebulosas espirais ainda geravam muita discordância até que, em 1920, promoveu-se um grande debate entre os dois principais representantes das duas ideias principais sobre o significado das nebulosas espirais (ou extragalácticas). De um lado Herber D. Curtis, que argumentava que cada uma destas nebulosas era um “universo-ilha” o qual era formado de estrelas e gás e que a Via Láctea era o universo-ilha ao qual pertencíamos e estávamos localizados em uma posição central. Do outro lado, Harlow Shapley argumentava que estas nebulosas eram nuvens de gás presentes no nosso Universo e que este era formado somente pela Via Láctea donde estávamos situados muito distantes do centro.

Ambas ideias provaram-se parcialmente incorretas por Edwin Hubble em meados da década 1920. Enquanto Shapley estava errado quanto à nossa posição na Galáxia por não ter levado em consideração a atenuação da luz causada pela poeira, Hubble observando estrelas na galáxia vizinha de Andrômeda, provou que a distância dela à Via Láctea era ainda maior do que a que o próprio Curtis acreditava ser o tamanho de nossa galáxia².

Nasce aí a astronomia extragaláctica. Desde seu princípio, realizaram-se *surveys* destes objetos, as galáxias, a fim de realizar classificações em função de sua morfologia, descobrir suas estruturas internas e sua evolução com o tempo. Pode-se dizer que o primeiro *survey* realizado nos moldes dos modernos *surveys* atuais foi realizado por Shane et al. (1959) e, desde então, a quantidade de dados não para de crescer.

Com as novas gerações de instrumentos impulsionada principalmente pelo advento dos detectores de estado sólido, temos observamos uma expansão gigantesca na taxa de aquisição de dados. Observações de diversos telescópios, tanto na terra quanto no espaço, estão gerando volumes de dados a uma taxa da ordem de *petabytes* por ano com uma tendência de crescimento exponencial tanto em quantidade quanto em complexidade (Way 2011).

No campo dos *mega-surveys* atuais, podemos nomear o *Sloan Digital Sky Survey* (SDSS) como o mais produtivo em quantidade de dados e publicações da última década. Para ter-se uma ideia do impacto deste na comunidade astronômica mundial, em 10 anos produziu-se com dados do SDSS mais de 3200 artigos, com mais de 135 mil citações (Driver 2011)! Não resta dúvida que o SDSS pode ser considerado referência quando falamos em levantamentos modernos. Em seus mais de oito anos de funcionamento,

²http://apod.nasa.gov/diamond_jubilee/debate20.html

obteve imagens e espectros de um quarto do céu, mapeando tridimensionalmente mais de 930.000 galáxias e 120.000 quasares, sem contar estrelas, cometas, aglomerados de estrelas e outras observações indiretas.

Isto é somente a ponta do *iceberg* de dados que nos aguarda. Outros *surveys* contribuirão para este aumento de dados já estão em fase de projeto ou tomando seus primeiros dados como, por exemplo, VST, VISTA, PanSTARRs, DES, LSST e Euclid (todos esses no ótico), WISE (infravermelho), XMM-XLL (raios X) e ASKAP e MeerKA (Driver 2011, rádio). Com esta quantidade literalmente astronômica de dados, gostaríamos de aprender ainda mais sobre a composição das galáxias.

1.1 O STARLIGHT e o mega-levantamento astronômico SDSS

Um exemplo do que se pode fazer com tamanha quantidade de dados de espectroscopia ótica como os que o SDSS nos proporciona é o código de síntese espectral STARLIGHT (Cid Fernandes et al. 2005), que nosso grupo desenvolveu e vem aprimorando nos últimos anos.

O STARLIGHT hoje é amplamente conhecido e utilizado nos quatro cantos do mundo. Ele ajusta espectros observados Ångström a Ångström a partir de combinações lineares de populações estelares simples (SSPs, do inglês *simple stellar populations*), cujos espectros vêm dos modelos de Bruzual & Charlot (2003). Resumidamente, o código de síntese espectral processa os dados observados (espectros) a um espaço de propriedades físicas com massas estelares, idades e metalicidades médias, extinção, dispersão de velocidades estelares e o próprio histórico de formação estelar (SFH, do inglês).

Os resultados obtidos da aplicação do STARLIGHT a 926.246 galáxias do *Data Release 7* do SDSS são excelentes, e têm sido utilizados em diversos estudos tanto por nosso grupo (Cid Fernandes et al. 2005; Mateus et al. 2006; Stasińska et al. 2006; Cid Fernandes et al. 2007; Mateus et al. 2007; Asari et al. 2007; Stasińska et al. 2008; Vale Asari et al. 2009; Cid Fernandes et al. 2010) quanto por outros grupos (Liang et al. 2007; Peebles et al. 2009) devido à publicidade tanto do programa de síntese quanto destas análises em um serviço de base de dados disponível online³.

³<http://www.starlight.ufsc.br/>

Outros projetos de códigos de ajuste espectral de galáxias que têm ajustes bem parecidos podem ser enumerados: MOPED (Panter et al. 2003, 2007), PLATEFIT (Tremonti et al. 2004), VESPA (Tojeiro et al. 2007), STECKMAP (Ocvirk et al. 2006), sedfit (Walcher et al. 2006), NBURSTS (Chilingarian et al. 2007) e ULySS (Koleva et al. 2009).

A partir do STARLIGHT e de seu completo banco de dados de propriedades físicas e linhas de emissão obtido da sua aplicação ao SDSS é que vamos basear este trabalho.

Para este trabalho, nos basearemos nos resultados obtidos da síntese de galáxias do SDSS utilizando o STARLIGHT.

1.2 O mega-levantamento J-PAS e este trabalho

O projeto J-PAS (*Javalambre Physics of the Accelerating Universe Astrophysical Survey*⁴) é uma colaboração Hispano-Brasileira da qual o Grupo de Astrofísica da UFSC faz parte juntamente com mais de 70 outros pesquisadores e que pretende, em 3 anos, mapear 8000 graus quadrados do céu em imagens a partir de 56 filtros de largura média de 125 Å. Espera-se que, ao final de suas atividades, este *survey* proporcione “espectros de baixa resolução” para cerca de 300 milhões de galáxias. Essa é uma grande quantidade de dados que merece ser explorada na sua totalidade.

O principal objetivo científico ao mapear tamanha área do céu em tantos filtros é a exploração das causas da aceleração da expansão do Universo, observada nas últimas décadas. A natureza desta componente que causaria esta aceleração, a energia escura (correspondente a 75% do Universo), atualmente é considerada a primeira das 25 “Grandes Perguntas” para os próximos 25 anos pela revista Science (Seife 2005). Com esta quantidade de filtros e as novas técnicas de obtenção de *redshifts* fotométricos, espera-se obter uma precisão de $\sigma_{z,\text{fotom.}}/(1+z) \approx 0,003$, ideal para o estudo preciso das oscilações acústicas bariônicas (Benítez et al. 2009).

Dada esta quantidade inédita de imagens em tantos filtros, há uma infinidade de outros casos científicos que se aproveitarão dos dados obtidos pelo JPAS.

Entre os vários benefícios indiretos do levantamento, podem ser listados:

- Mapeamento de *redshifts* de quasares e suas vizinhanças correlacionando sistemas

⁴<http://j-pas.org/>

absorvedores com as galáxias observadas na linha de visada, fornecendo informações sobre distribuição de gás ao redor das galáxias;

- Determinação de temperaturas efetivas, gravidade e metalicidade em estrelas, principalmente no halo Galáctico;
- Detecção e classificação em famílias de asteroides através de suas cores;
- Estudo de propriedades físicas de galáxias, tanto de suas estrelas como do gás.

Nesta dissertação, investigaremos quão precisamente podemos tratar o último item. Quando olhamos os dados de experimentos como o JPAS, gostaríamos que os dados nos contassem informações sobre as populações estelares das galáxias. Além disto, gostaríamos também de obter informações sobre suas linhas de emissão.

Uma das formas de abordar o problema (figura 1.1), seria realizar ajustes de populações estelares, como é feito pelo STARLIGHT: Utilizar-se de modelos de síntese evolutiva de populações estelares para inferir a partir de ajustes espectrais sobre a História de Formação Estelar (SFH, do inglês *Star Formation History*) e medir as linhas de emissão no espectro residual.

Este é um método. Porém, ao realizar tal tipo de ajustes, incidiríamos em sérias limitações proporcionais às limitações em resolução espectral do JPAS: $\Delta\lambda \approx 100\text{\AA}$ ou $R = \frac{\lambda}{\Delta\lambda} \approx 50$. O que o STARLIGHT e programas similares fazem é decodificar informação sobre populações estelares na forma do contínuo e linhas de absorção. A forma do contínuo será bem traçada pelo JPAS, porém as linhas de absorção não. Esta perda de informação deve acarretar em uma baixa da precisão na medida das propriedades. Medidas de linhas de emissão no espectro residual do JPAS seriam praticamente impossíveis com dada resolução espectral, como também separar algumas linhas de emissão, como por exemplo $[\text{N II}]\lambda 6584$ de $\text{H}\alpha$, inviabilizando o cálculo de razões de linhas de emissão, importantes para muitas aplicações.

Exploramos aqui uma alternativa que pode minimizar estas dificuldades ou, pelo menos, nos dar a mesma quantidade de informações que uma abordagem pelo ajuste de modelos similar à do STARLIGHT nos proporcionaria.

Nossa estratégia é utilizar o SDSS como uma grade referência para comparações ao invés de utilizar uma grade composta por modelos de populações estelares (como as de Bruzual & Charlot (2003) utilizadas pelo STARLIGHT). Assim, ao invés de compararmos uma galáxia (gás + estrelas + poeira + ...) com modelos populações estelares,

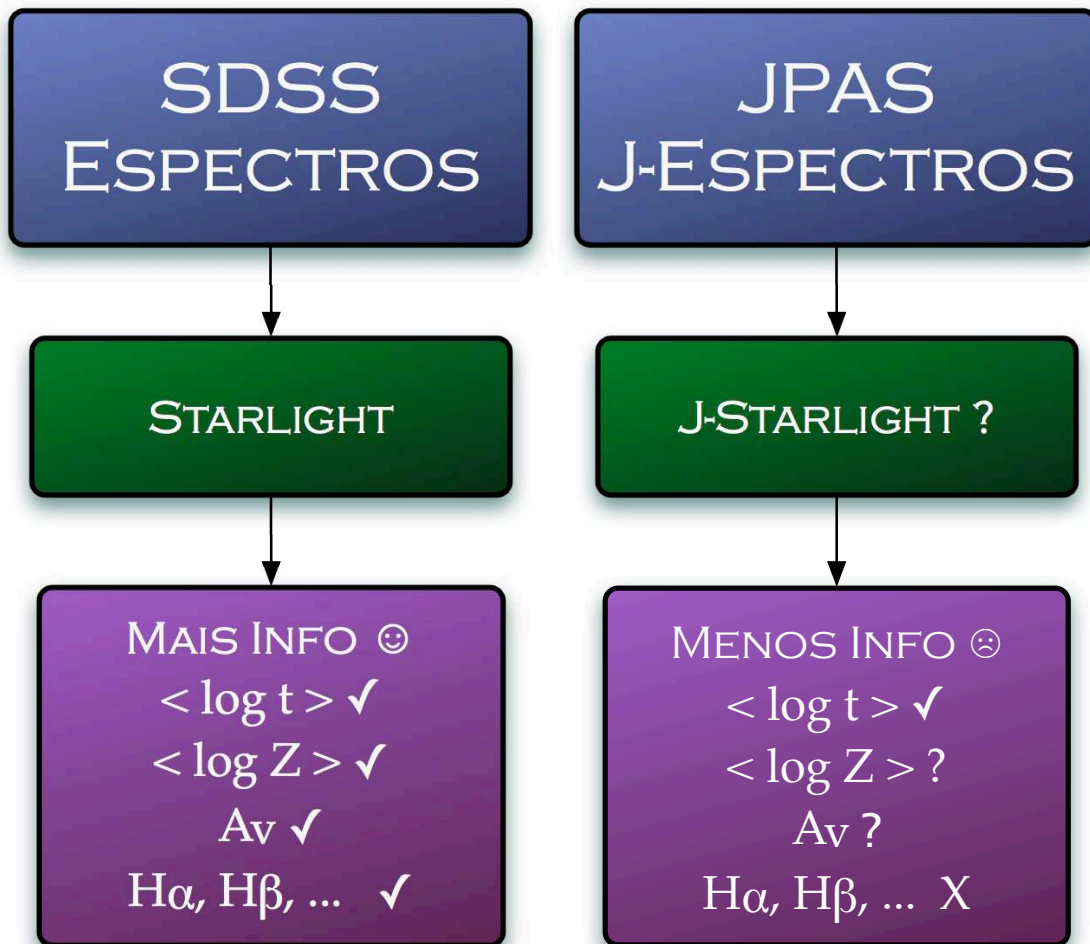


Figura 1.1: Esquema gráfico de um modo de análise para os dados JPAS. Nosso método para síntese usual consiste em, dado um espectro de uma galáxia obtido pelo SDSS, encontrar a combinação linear de modelos de populações estelares que melhor se ajustam (STARLIGHT). Com isto, obtemos todos os parâmetros físicos da galáxia e do espectro residual as medidas da linha de emissão. Isto poderia ser feito da mesma forma para os dados do JPAS, porém, devido à sua baixa resolução, estes seriam recuperados com muito menor precisão.

comparamos galáxias do JPAS com galáxias SDSS e, de uma maneira simples, assumimos que se duas galáxias possuem J-espectros (pontos fotométricos do JPAS) iguais (ou muito semelhantes), estas possuem propriedades físicas e linhas de emissão iguais (ou muito parecidas). A princípio, ao comparar o espectro de uma galáxia do JPAS com uma do SDSS, poderia-se inferir sobre propriedades que necessitariam de uma resolução espectral muito maior que a disponível, como a separação entre [N II] e $H\alpha$ já mencionada.

Para verificar a utilidade do método proposto, gostaríamos saber qual a precisão para as medidas das seguintes propriedades:

- **Populações Estelares:** Idades estelares, extinção, metalicidades estelares, razão massa luminosidade e outras propriedades físicas;
- **Linhas de emissão:** Larguras equivalentes de linhas de emissão (como $H\alpha$, [O II], ...), razões entre linhas de emissão (como [N II]/ $H\alpha$, [O III]/ $H\beta$, ...) e outras propriedades observadas quando temos espectros de alta resolução.

E, ainda, como esta precisão varia em função de:

- Escolha da amostra de comparação;
- Escolha do esquema de comparação (pesos e priores);
- *Redshift*;
- Relação sinal-ruído S/N ;

Resumindo, a questão fundamental desta dissertação seria: O que se pode aprender sobre galáxias a partir de *surveys* fotométricos como o JPAS? Tentaremos responder esta questão e algumas outras que surgirão ao longo de nosso estudo nos próximos capítulos.

Capítulo 2

Do SDSS ao J-PAS

RESUMO: Neste capítulo apresentamos brevemente as características dos levantamentos astronômicos SDSS e J-PAS (seções 2.1 e 2.2). Em seguida, apresentamos os cálculos necessários para converter um espectro como o do SDSS para um conjunto de filtros como os do J-PAS (seção 2.3) e os erros associados a cada ponto (seção 2.4). Por fim, na seção 2.5, mostramos como isto foi feito na prática e definimos o elemento básico que será trabalhado nos próximos capítulos: o J-espectro.

2.1 O megalevantamento SDSS

O SDSS (York et al. 2000) é um levantamento de dados astronômicos que, sem sombra de dúvidas, revolucionou a maneira de se fazer astronomia. Foi o primeiro grande levantamento que utilizou *Charged Coupled Devices* (CCDs) ao invés de placas fotográficas para a obtenção de dados fotométricos e espectroscópicos. Além da quantidade jamais vista de dados, o que leva este *survey* a destacadamente ser um sucesso em comparação aos seus antecessores é a facilidade proporcionada no acesso aos seus dados através de uma interface de consulta¹.

O levantamento utiliza um telescópio exclusivo de 2,5m localizado no *Apache Point Observatory* (Novo México, EUA). Este é equipado com uma câmera CCD, dedicada à fotografia do céu em 5 bandas fotométricas (u , g , r , i e z), e dois espectrógrafos multifibras com capacidade de realizar espectroscopia de até 640 objetos simultaneamente com uma resolução média de 3\AA . No *Data-Release 7* (DR7) foram observadas espectro-

¹<http://cas.sdss.org/astrodr7/en/>

scopicamente 929555 galáxias espalhadas em uma região do céu correspondente a 9380 graus quadrados (figura 2.1) em uma faixa espectral que varia entre 3800 e 9200 Å e magnitude na banda r menor que 17,77.

Os campos são primeiramente observados fotometricamente a fim de detectar em quais objetos será realizada a espectroscopia. Então, placas de alumínio (figura 2.2) são perfuradas com as posições exatas onde as fibras deverão estar localizadas. As fibras são conectadas ao espectrógrafo e os espectros são coletados por um CCD.

Para este trabalho, utilizamos os espectros de 926246 destas galáxias reamostrados em 1Å e no referencial de repouso utilizando os *redshifts* (z) determinados pela pipeline do SDSS. Do SDSS, foram extraídos os fluxos para cada comprimento de onda, os erros associados a cada um destes fluxos e uma *flag* que indica regiões com problemas no espectro. Exemplos de espectros do SDSS são mostrados na figura 2.5.

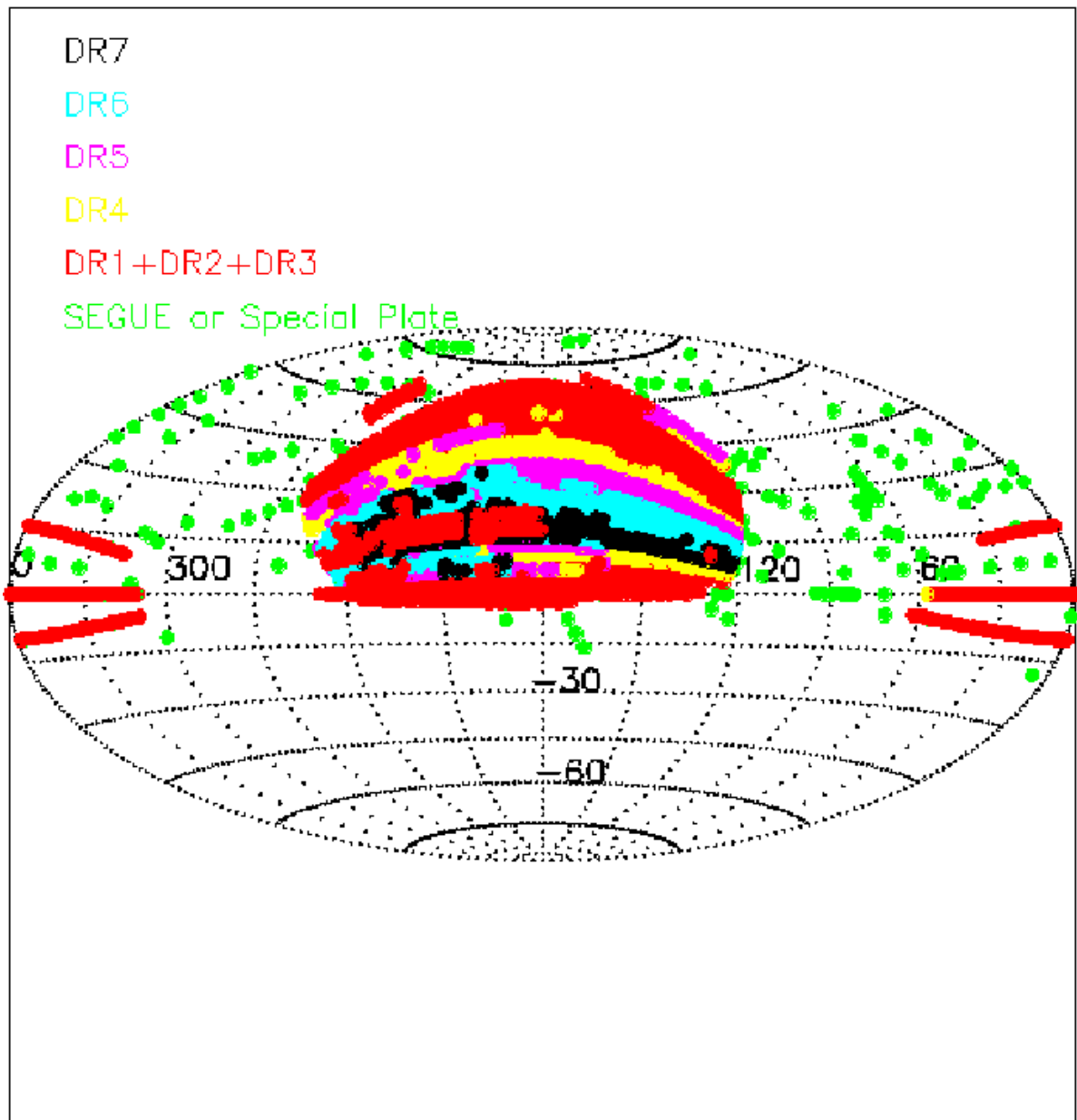


Figura 2.1: Projeção equatorial das regiões cobertas pelo levantamento espectral do SDSS para seus diversos Data-Releases. Neste trabalho, os dados utilizados foram os do DR7, que engloba seus predecessores. Figura extraída de <http://www.sdss.org/>

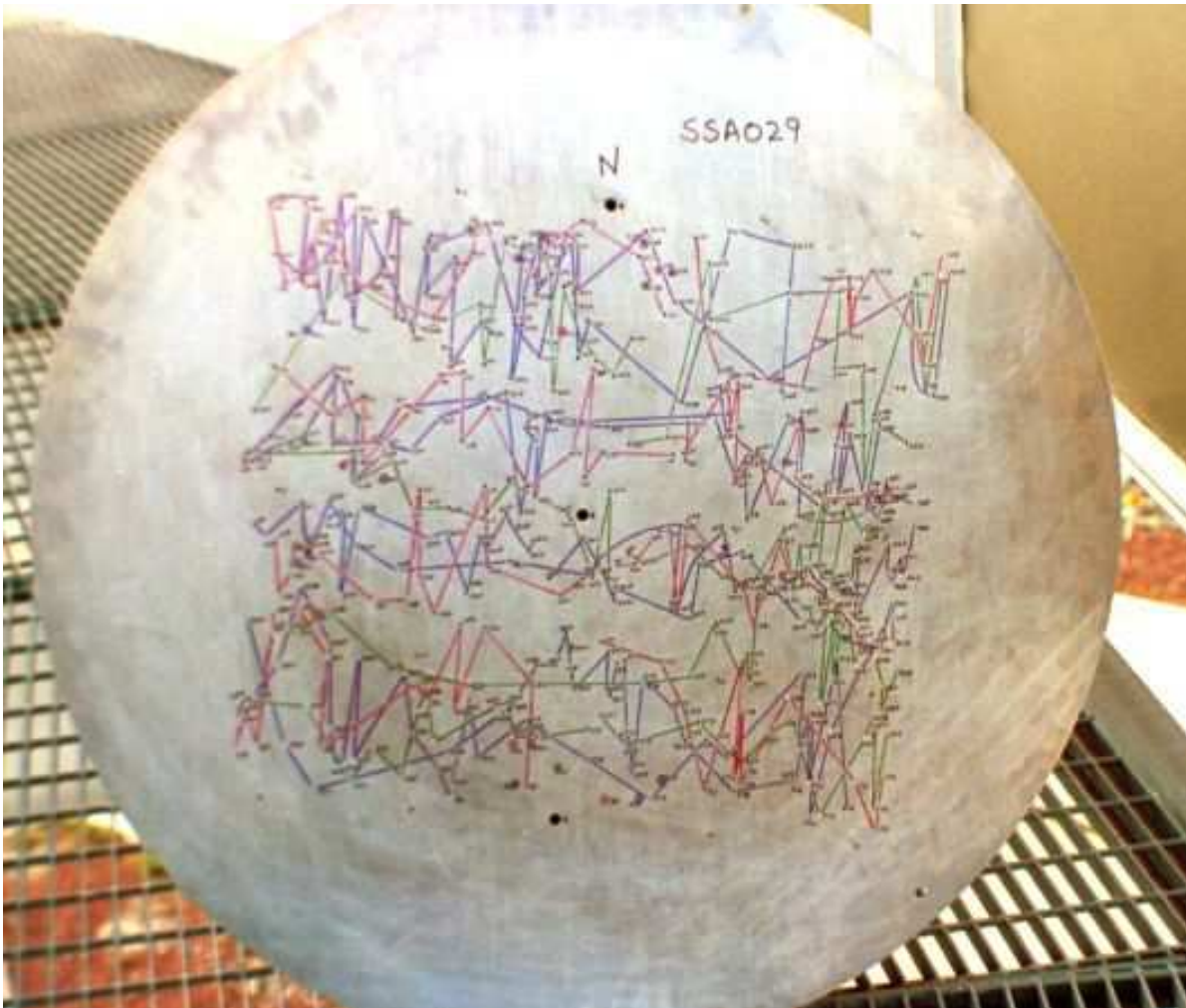


Figura 2.2: Exemplo de um prato de fibras do SDSS. Após fotografar o campo com um CCD, os pratos são marcados por um software e furados a fim de localizar as 640 fibras do espectrógrafo nos lugares correspondentes aos objetos que deseja-se analisar. Figura extraída de <http://www.sdss.org/>

2.2 O megalevanteamento J-PAS

O J-PAS (*Javalambre Physics of Accelerating Universe Astrophysical Survey*²) é um projeto fruto de uma colaboração entre instituições brasileiras e espanholas que deverá começar a produzir dados em 2012. Utilizando-se um telescópio de 2,52m, instalado no recém-criado Observatório Astronômico Javalambre em Teruel (OAJ) na Espanha, irá fotografar com um mosaico de 14 CCDs (figura 2.3) 8000 graus quadrados do céu em 56 filtros de largura em comprimento de onda muito estreitos (da ordem de 125 Å, figura 2.4).

Como o produto das observações são imagens e não espectros propriamente, perde-se na resolução espectral, quando comparada aos 3Å do SDSS, porém ganha-se resolução espacial. Ao invés de coletar para cada galáxia apenas a luz proveniente dos 3 segundos de arco centrais, teremos a informação *pixel a pixel* de cada galáxia onde cada pixel corresponde a aproximadamente 0,23 segundos de arco. Ainda, nestes 8000 graus, espera-se fotografar 300 milhões de galáxias, ou seja, 300 vezes mais que no SDSS. Isto o torna um grande laboratório para estudos estatísticos das propriedades físicas e observacionais de galáxias, como propomos neste trabalho.

Embora o escopo geral do projeto já esteja bem definido, vários aspectos técnicos ainda estão em estudo, como o perfil dos filtros, a estratégia de observação e aspectos da calibração e redução. Tais “detalhes”, no entanto, não afetam de maneira importante o estudo aqui apresentado.

Para simular dados do JPAS e investigar quanta informação poderemos extrair deles utilizaremos espectros do SDSS.

²<http://j-pas.org/>

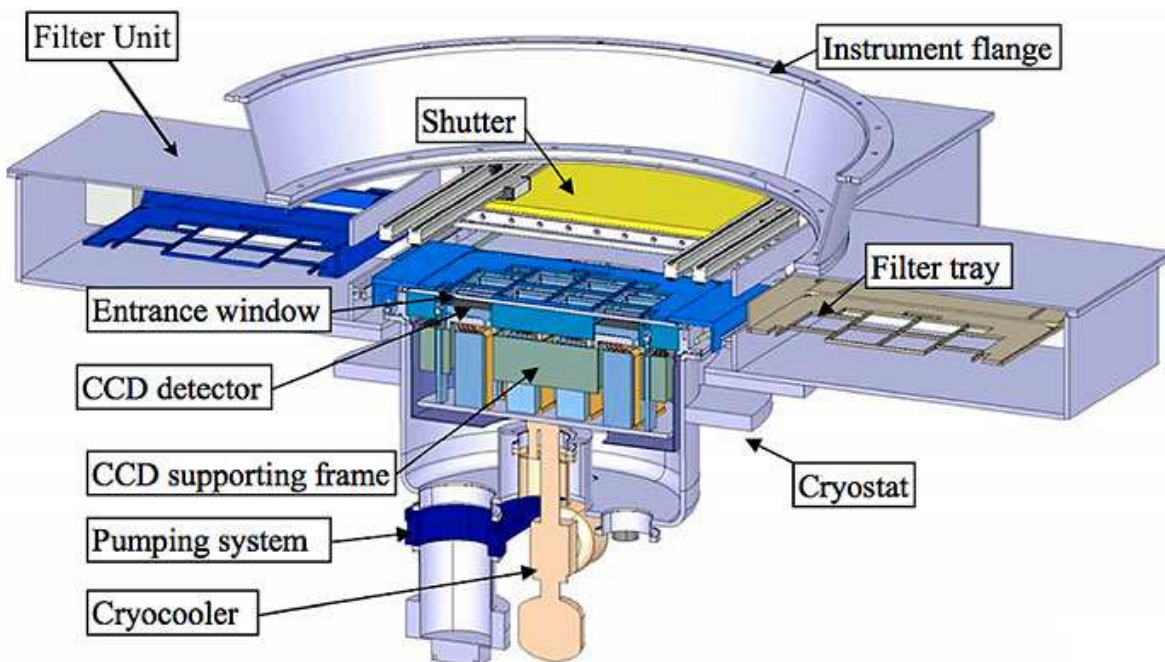


Figura 2.3: Desenho conceitual do mosaico de CCDs que equipará o telescópio de 2,52m no OAJ. Figura extraída de <http://j-pas.org/>.

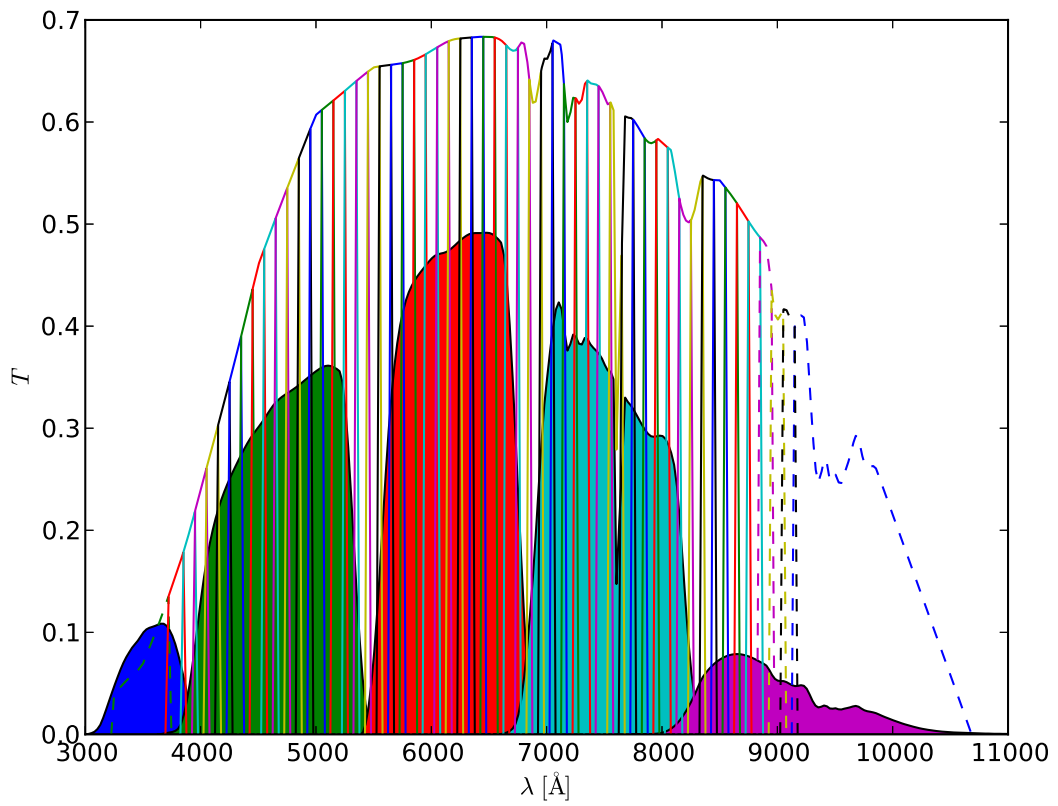


Figura 2.4: Curvas de transmissividade utilizadas neste trabalho. Cada curva estreita colorida representa um dos 56 filtros de $\approx 125 \text{\AA}$ do J-PAS. Para este trabalho, devido à faixa de cobertura espectral do SDSS, foram excluídos os 4 filtros marcados em linha pontilhada. As cinco curvas abaixo da do JPAS representam as curvas u , g , r , i e z do SDSS.

2.3 Conversão de Espectros do SDSS para o J-PAS

Para realizarmos a conversão de um espectro do SDSS (e os erros associados a este) a dados fotométricos do J-PAS, utilizamos a seguinte relação:

$$J_l = \frac{\int O_\lambda T_{l,\lambda} d\lambda}{\int T_{l,\lambda} d\lambda} = \sum_\lambda O_\lambda t_{l,\lambda} \Delta\lambda \quad (2.1)$$

onde O_λ é o espectro observado no SDSS, $T_{l,\lambda}$ é a curva de transmissividade para um determinado filtro l e $t_{l,\lambda}$ é dado pela relação:

$$t_{l,\lambda} \equiv \frac{T_{l,\lambda}}{\sum_\lambda T_{l,\lambda} \Delta\lambda} \quad (2.2)$$

tal que J_l é o fluxo em um dado filtro l . Definidos dessa maneira, J_l e O_λ têm a mesma unidade, [$\text{erg s}^{-1} \text{cm}^{-2} \text{Å}^{-1}$]. Exemplos desta conversão são mostrados na figura 2.5.

Para calcularmos os espectros acima, os espectros do SDSS foram trazidos para o referencial de repouso e corrigidos pelo avermelhamento Galáctico. Para simplificação, utilizaremos todos espectros no referencial de repouso ($z = 0$). Esperamos que a partir dos *redshifts* fotométricos calculados pelos métodos tradicionais (Benítez 2000), o ajuste deste termo seja desnecessário.

2.4 Espectros de erros para o espectro J-PAS

Para calcularmos o erro correspondente ao fluxo J_l em cada filtro, utilizamos o erro no espectro observado no SDSS. A matemática desta transformação é a seguinte:

Temos nosso espectro observado O_λ como uma soma do espectro intrínseco da galáxia O_λ^0 mais o ruído n_λ :

$$O_\lambda = O_\lambda^0 + n_\lambda \quad (2.3)$$

e, portanto,

$$O_\lambda^2 = (O_\lambda^0 + n_\lambda)^2 = O_\lambda^{0^2} + 2O_\lambda^0 n_\lambda + n_\lambda^2 \quad (2.4)$$

Assumindo que o valor esperado do ruído seja nulo, $E(n_\lambda) = 0$, e que a variância do ruído de O_λ , $\sigma^2(n_\lambda)$, seja conhecida (o SDSS provê esta informação), podemos calcular a variância no J-espectro a partir da relação:

$$\sigma^2(J_l) = E(J_l^2) - E^2(J_l) \quad (2.5)$$

onde $E(x)$ é o valor esperado de x .

Podemos, com a equação 2.1, calcular o valor esperado de J_l :

$$\begin{aligned} E(J_l) &= E\left(\sum_\lambda O_\lambda t_{l,\lambda} \Delta\lambda\right) \\ &= E\left(\sum_\lambda (O_\lambda^0 + n_\lambda) t_{l,\lambda} \Delta\lambda\right) \\ &= E\left(\sum_\lambda O_\lambda^0 t_{l,\lambda} \Delta\lambda\right) + E\left(\sum_\lambda n_\lambda t_{l,\lambda} \Delta\lambda\right) \\ &= J_l^0 + \sum_\lambda E(n_\lambda) t_{l,\lambda} \Delta\lambda = J_l^0 \end{aligned} \quad (2.6)$$

onde usamos o fato de que $E(n_\lambda) = 0$. De modo similar, podemos calcular o valor esperado para J_l^2 :

$$\begin{aligned} E(J_l^2) &= E\left(\sum_i (O_i^0 + n_i) t_{l,i} \Delta\lambda \sum_j (O_j^0 + n_j) t_{l,j} \Delta\lambda\right) \\ &= E\left(\left(\sum_i O_i^0 t_{l,i} \Delta\lambda + \sum_i n_i t_{l,i} \Delta\lambda\right) \left(\sum_j O_j^0 t_{l,j} \Delta\lambda + \sum_j n_j t_{l,j} \Delta\lambda\right)\right) \\ &= J_l^{0^2} + 2J_l^0 E\left(\sum_i n_i t_{l,i} \Delta\lambda\right) + E\left(\sum_i n_i t_{l,i} \Delta\lambda \sum_j n_j t_{l,j} \Delta\lambda\right) \\ &= J_l^{0^2} + E\left(\sum_i n_i t_{l,i} \Delta\lambda \sum_j n_j t_{l,j} \Delta\lambda\right) \end{aligned} \quad (2.7)$$

O último termo pode ser dividido em duas somas. Uma para $i = j$ e outra para $i \neq j$. Assim,

$$\begin{aligned} E\left(\sum_i n_i t_{l,i} \Delta\lambda \sum_j n_j t_{l,j} \Delta\lambda\right) &= E\left(\sum_i n_i^2 t_{l,i}^2 \Delta\lambda^2 + \sum_i n_i t_{l,i} \Delta\lambda \sum_{j \neq i} n_j t_{l,j} \Delta\lambda\right) \\ &= \sum_i E(n_i^2) t_{l,i}^2 \Delta\lambda^2 + \sum_i \sum_{j \neq i} E(n_i n_j) t_{l,i} t_{l,j} \Delta\lambda^2 \quad (2.8) \end{aligned}$$

Assumindo que os ruídos n_i e n_j não sejam correlacionados na janela do filtro, n_i é independente de n_j , então o segundo termo da equação 2.8 é igual a zero. Assim,

$$E(J_l^2) = J_l^0{}^2 + \sum_i E(n_i^2) t_{l,i}^2 \Delta\lambda^2 \quad (2.9)$$

Substituindo as equações 2.6 e 2.9 em 2.5, podemos encontrar uma expressão para $\sigma^2(J_l)$:

$$\sigma^2(J_l) = E(J_l^2) - E(J_l)^2 = (J_l^0{}^2 + \sigma^2(J_l)) - J_l^2 \quad (2.10)$$

$$\begin{aligned} \sigma^2(J_l) &= N_{\lambda,l} \langle \sigma(n_\lambda)^2 t_{l,\lambda}^2 \Delta\lambda^2 \rangle \\ &= N_{\lambda,l} \Delta\lambda \langle \sigma(n_\lambda)^2 t_{l,\lambda}^2 \rangle \Delta\lambda \\ &= \Lambda_l \langle \sigma(n_\lambda)^2 t_{l,\lambda}^2 \rangle \Delta\lambda \quad (2.11) \end{aligned}$$

onde $N_{\lambda,l}$ é o número de *pixels* espectrais no filtro l e $\Lambda_l \equiv N_{\lambda,l} \Delta\lambda$ é o tamanho da janela espectral do filtro.

Para termos uma ideia aproximada de como o erro em J_l depende do erro em O_λ podemos considerar um conjunto de filtros retangulares, para os quais a função de transmissividade é simplesmente $t_{l,\lambda} = \frac{1}{\Lambda_l}$. Neste caso temos:

$$\begin{aligned}\sigma^2(J_l) &= \Lambda_l \Delta\lambda \langle \sigma^2(n_\lambda) t_{l,\lambda}^2 \rangle \\ &= \frac{\Delta\lambda}{\Lambda_l} \langle \sigma^2(n_\lambda) \rangle\end{aligned}$$

Substituindo os valores típicos $\Lambda_l^{\text{JPAS}} \approx 100 \text{ \AA}$ e $\Delta\lambda^{\text{SDSS}} \approx 3 \text{ \AA}$, temos:

$$\sigma(J_l) = \sqrt{\frac{3}{100} \langle \sigma^2(n_\lambda) \rangle}$$

Esperamos, então, que nossos erros para cada filtro do JPAS sejam da ordem de $\sqrt{\frac{100}{3}} \approx 6$ vezes menor que em cada *pixel* do SDSS na mesma faixa espectral.

2.5 O J-espectro

Partindo das equações 2.1 e 2.11, podemos degradar os espectros do SDSS para simular os J-espectros. Porém, para converter estes espectros, temos que resolver a situação de quando não possuímos um ou mais valores para O_λ . Isto ocorre, por exemplo, devido a problemas com o detector do SDSS ou na remoção do céu. Neste caso, adotamos a seguinte receita em cada filtro do JPAS:

- Caso o número de pontos com problemas seja superior a 50% do número de comprimentos de onda no filtro, este J-pixel é descartado nas análises.
- Caso o número de pontos com problemas seja inferior ou igual a 50% do número de comprimentos de onda no filtro, usamos o fluxo observado do SDSS para os pontos sem problemas e o fluxo sintético do STARLIGHT para os pontos com problemas. No cálculo do J-erro, pontos com problemas são ignorados na equação 2.11.

Assim, transformamos todo os espectros do SDSS-DR7 para J-espectros e os utilizamos como base para este trabalho.

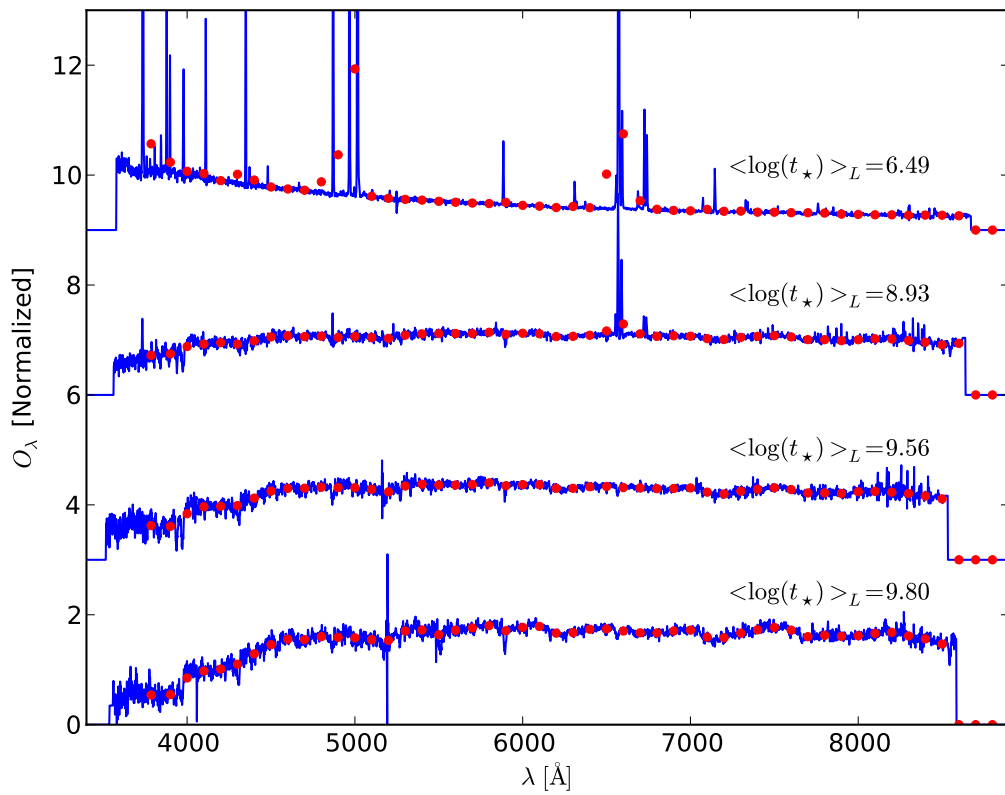


Figura 2.5: Exemplos de J-espectros normalizados distribuídos em $\langle \log t_* \rangle_L$ que é uma medida da idade média em anos (ponderada pela luz) das populações estelares em uma galáxia. Em azul o espectro observado pelo SDSS e em vermelho o J-espectro correspondente.

Capítulo 3

Método

RESUMO: *Este capítulo apresenta o método de casamento espectral adotado para extrair informação dos J-espectros de forma indireta, utilizando resultados obtidos através da aplicação do STARLIGHT no SDSS. Começamos com uma descrição da filosofia geral do método (seção 3.1) e em seguida definimos a notação usada para os ingredientes básicos envolvidos no casamento (seção 3.2). Na seção 3.3 apresentamos a figura de mérito utilizada para o casamento espectral bem como a receita adotada para dar pesos aos pixels de cada J-espectro. Apresentamos, então, alguns exemplos de casamentos espectrais na seção 3.4 e finalizamos com a seção 3.5 onde apresentamos a abordagem probabilística utilizada para inferir propriedades físicas e observacionais a partir dos resultados dos casamentos espectrais.*

3.1 Introdução

O SDSS sem dúvida foi divisor de águas na astronomia como um todo tanto pela quantidade de dados como em respeito à forma que estes foram reduzidos, organizados e disponibilizados. Nosso grupo, com a aplicação do STARLIGHT para 926246 galáxias, pôde ao longo dos últimos anos construir (e disponibilizar em <http://starlight.ufsc.br/>) um banco de dados para uma série de propriedades físicas (massa estelar, extinção, idade média e outros) e medidas de fluxos e larguras equivalentes de linhas de emissão com grande precisão (Asari et al. 2007).

No capítulo 1 vimos quão bem este maquinário funciona e, na figura 1.1, especulamos que talvez seja possível para os dados do J-PAS seguir uma rota de trabalho similar à

aplicada por nós ao estudo do SDSS. Porém, devido à baixa resolução espectral dos dados, esta abordagem direta aos dados nos traria muito menos informação. Por exemplo, é intuitivamente óbvio que não poderemos extrair histórias de formação estelar de J-espectros com o mesmo grau de detalhe que fazemos no SDSS. E, muito menos, medidas diretas de linhas de emissão. Será impossível, por exemplo, separar as linhas de $[N II]$ e $H\alpha$ em J-espectros, o que acarreta sérias limitações em termos de muitas aplicações astrofísicas.

Assim, a ideia chave deste trabalho é tomar um **atalho** que nos remeta às propriedades recuperadas do SDSS partindo dos dados do J-PAS a partir do casamento espectral, como esquematizado na figura 3.1.

Assumimos que se duas galáxias possuem fluxos em 56 filtros estreitos iguais ou muito parecidos elas devem ter seus espectros em média-alta resolução também muito parecidos e extrapolamos isto para que, se os espectros são parecidos, suas propriedades físicas e linhas de emissão também serão.

Esta ideia é uma variação de uma metodologia muito utilizada hoje em dia como, por exemplo, por vários autores que desenvolveram códigos que comparam alguns índices espectrais a conjuntos de modelos gerados a partir de cálculos de síntese evolutiva (Bruzual & Charlot 2003) como fazem Gallazzi et al. (2005). No nosso caso, não compararemos a modelos mas sim a espectros de fato observados, o que nos traz uma série de vantagens, como a possibilidade da medida indireta de linhas de emissão que veremos nos próximos capítulos.

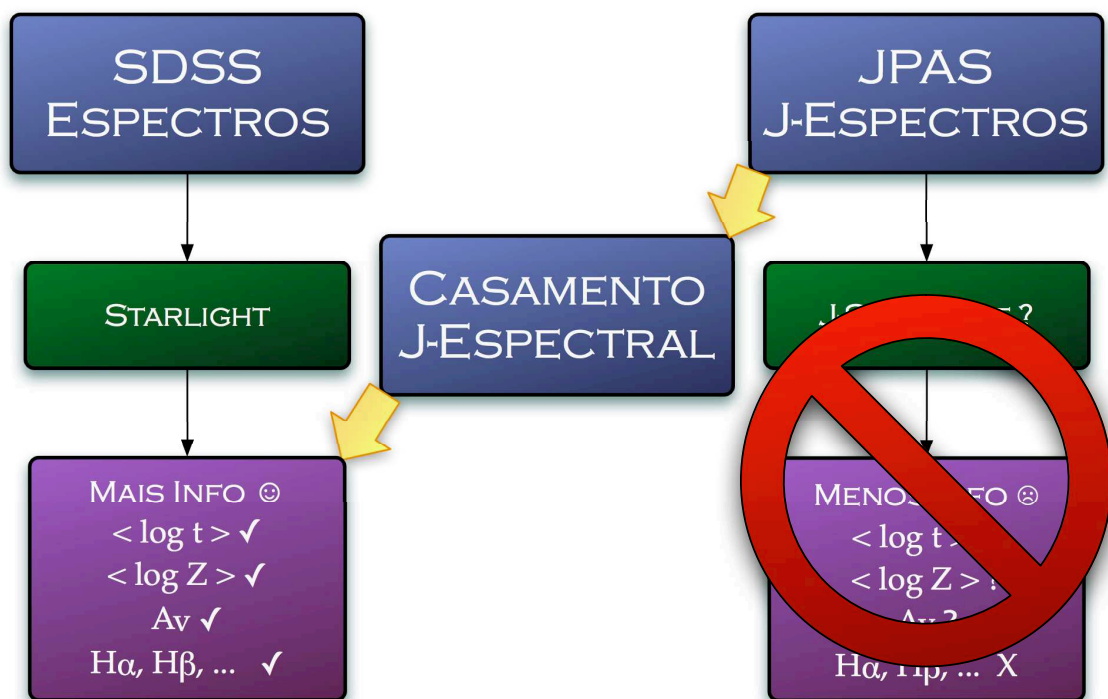


Figura 3.1: Esquema gráfico do método proposto para abordar o problema da baixa resolução. Ao invés de adaptar o STARLIGHT para os dados do J-PAS, propomos o casamento espectral para obter propriedades observadas somente em uma resolução maior.

3.2 Definições

Os ingredientes básicos desta análise são um J-espectro observado O e seus erros associados, uma base B de J-espectros de referência para o casamento espectral e um conjunto de propriedades p a serem estimadas.

A fim de facilitar a exposição do método nas próximas seções, é didático definirmos uma notação para estes ingredientes da seguinte forma:

- $O_{l,i}$: É o fluxo no filtro l de um **objeto** i de uma amostra de objetos. Estes simularão galáxias observadas pelo J-PAS seguindo a transformação da seção 2.3.
- $B_{l,j}$: É o fluxo no filtro l de um elemento j de uma **base** de espectros para referência. Seguem a mesma transformação.
- $\sigma(B_{l,j})$ e $\sigma(O_{l,i})$: Erros associados aos J-fluxos em cada elemento da base e dos objetos conforme definidos na seção 2.4.
- $M_{l,i,j}$: É o fluxo no filtro l de um Modelo para o objeto i construído a partir do elemento j da base. O modelo é dado por $M_{l,i,j} = a_{i,j}B_{l,j}$, onde $a_{i,j}$ é um fator de escala a ser otimizado.

3.3 Casamento Espectral: Figura de mérito e pesos

Para o casamento entre os pares de J-espectros calculamos o $\chi^2_{i,j}$ na sua forma tradicional:

$$\chi^2_{i,j} = \sum_l (O_{l,i} - a_{i,j}B_{l,j})^2 w_{l,i,j}^2 \quad (3.1)$$

onde o fator de escala $a_{i,j}$ é obtido através da minimização do χ^2 pela expressão:

$$\frac{\partial \chi^2}{\partial a_{i,j}} = \frac{\partial \sum_l (O_{l,i} - a_{i,j}B_{l,j})^2 w_{l,i,j}^2}{\partial a_{i,j}} = 0$$

o que nos dá:

$$a_{i,j} = \frac{\sum_l O_{l,i} B_{l,j} w_{l,i,j}^2}{\sum_l B_{l,j}^2 w_{l,i,j}^2} \quad (3.2)$$

Há várias maneiras para a escolha dos pesos $w_{l,i,j}$, dependendo de quais características gostaríamos de dar mais importância no casamento. A mais tradicional seria,

$$w_{l,i,j}^2 = \frac{1}{\sigma^2(O_{l,i}) + a_{i,j}^2 \sigma^2(B_{l,i})} \quad (3.3)$$

onde as incertezas tanto em $O_{l,i}$ quanto em $M_{l,i}$ são consideradas.

Todavia, preferimos uma versão levemente diferente, dada por:

$$w_{l,i,j}^2 = \left(\frac{\langle O_{l,i}^2 \rangle}{O_{l,i}^2} \right) \left(\frac{1}{\sigma^2(O_{l,i}) + a_{i,j}^2 \sigma^2(B_{l,i})} \right) \quad (3.4)$$

onde o segundo termo é dado pela equação 3.3 e o denominador do primeiro termo é tal que dá um peso igual às regiões com maior e menor fluxos. O numerador deste foi inserido a fim de manter a adimensionalidade do peso.

Para entender melhor a motivação para a introdução deste termo *ad-hoc*, note que ele faz com que a diferença quadrática entre $O_{l,i}$ e o modelo $M_{l,j}$ em um filtro l seja proporcional a $[(O_{l,i} - M_{l,j})/O_{l,i}]^2$. Desta maneira, uma diferença no pico de uma linha de emissão (alto $O_{l,i}$) e no fundo de uma linha de absorção (baixo $O_{l,i}$) contribuem igualmente ao χ^2 (afora, é claro, os outros fatores envolvidos no segundo termo em 3.4). Julgamos esta uma estratégia mais adequada, mas cabe notar que esta decisão no fundo não é tão relevante, uma vez que resultados muito parecidos são obtidos seguindo a receita padrão da equação 3.3.

O leitor mais atento percebe que o peso dado pela equação 3.4 depende de $a_{i,j}$, o qual é dado pela equação 3.2 que, por sua vez, é dependente de $w_{l,i,j}$. Para resolver este fato, no cálculo do χ^2 foram realizadas iterações no cálculo de $a_{i,j}$ e $w_{l,i,j}$ até houvesse convergência. Em geral, não foram necessárias mais do que 3 ou 4 iterações para que o critério de convergência por nós estabelecido ($\Delta a_{i,j} < 10^{-5}$) fosse satisfeito.

Ao fim deste processo temos para cada objeto O_i um vetor $\chi_{i,j}^2$, um vetor $N_{i,j}$ dado pelo número de *pixels* considerados bons no J-espectro do objeto O_i e do j -ésimo elemento de base B_j simultaneamente e um vetor com os fatores de escala $a_{i,j}$ que melhor ajusta o objeto O_i ao elemento de base B_j . Com isso, calculamos

$$\chi_{i,j,\text{RED}}^2 = \frac{1}{N_{i,j}} \sum_l (O_{l,i} - a_{i,j} B_{l,j})^2 w_{l,i,j}^2 \quad (3.5)$$

que é chamado de χ^2 reduzido.

3.4 Casamento Espectral: Exemplos

A versão mais simples do casamento espectral é o pareamento de galáxias. Nele, pegamos o elemento j da base que tem o menor χ^2 em relação ao objeto i e supomos que a galáxia j é idêntica à i . Dizemos então, em uma aproximação grosseira, que todas as propriedades de i são idênticas às propriedades de j . Os exemplos de pares mostrados nas figuras 3.2 e 3.3 sugerem que esta seria uma estratégia promissora.

Todavia, quando olhamos além do menor χ^2 , observamos rapidamente que o pareamento não é necessariamente o melhor método já que, quando ordenamos por χ^2 e comparamos os espectros (figuras 3.4 e 3.5), constatamos que na própria base existem galáxias muito semelhantes e com propriedades muitas vezes diferentes entre si.

Ao observarmos tamanha semelhança entre os espectros que compõem os melhores ajustes nos veio a ideia de, ao invés de escolher um único par objeto-base para definir as propriedades de determinado objeto, dar um enfoque estatístico onde todas os elementos da base são levados em conta proporcionalmente à semelhança com o J-espectro observado. Decidimos então a dar um enfoque análogo ao adotado por Kauffmann et al. (2003), Brinchmann et al. (2004) e Gallazzi et al. (2005) exceto que, no caso deles, a comparação é feita a partir de modelos teóricos enquanto a nossa é eminentemente empírica.

Esta técnica nos permite medir muita coisa, em particular, coisas que não conseguiríamos medir de forma direta no J-espectro, como [N II] e H α separadamente, o que nos permite inferir sobre medidas de larguras equivalentes e razões de linhas de emissão

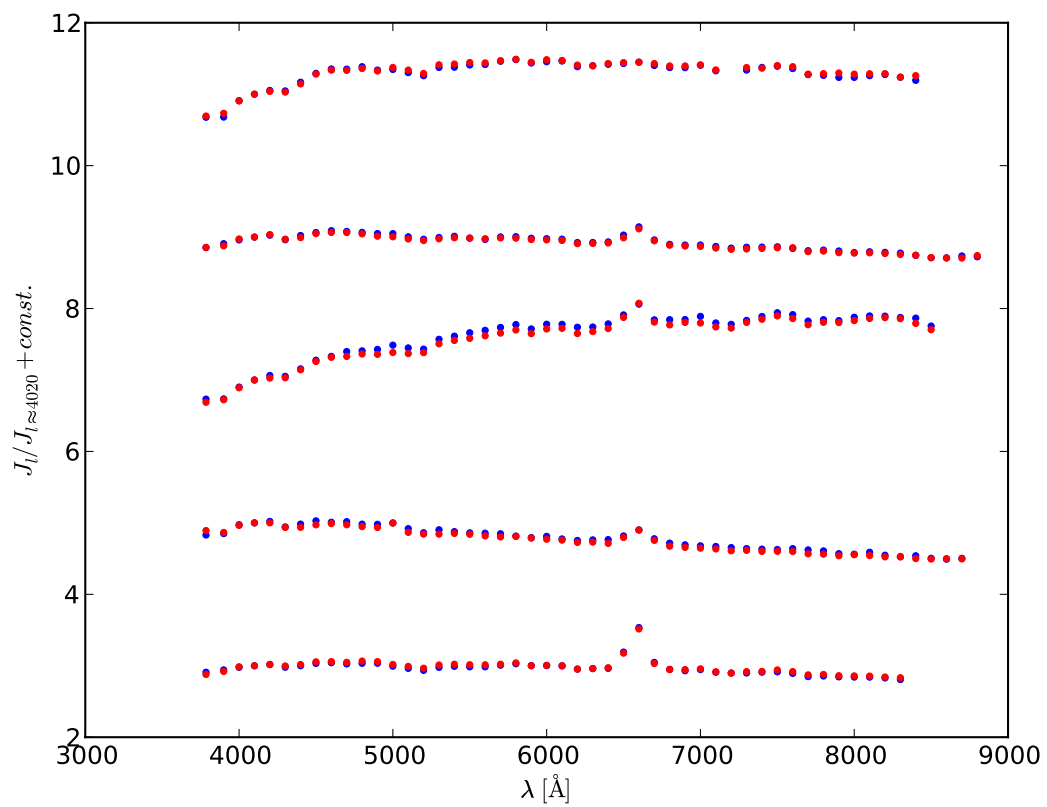


Figura 3.2: Quatro exemplos de J-espectros observados (em azul) e seus primeiros melhores ajustes (em vermelho).

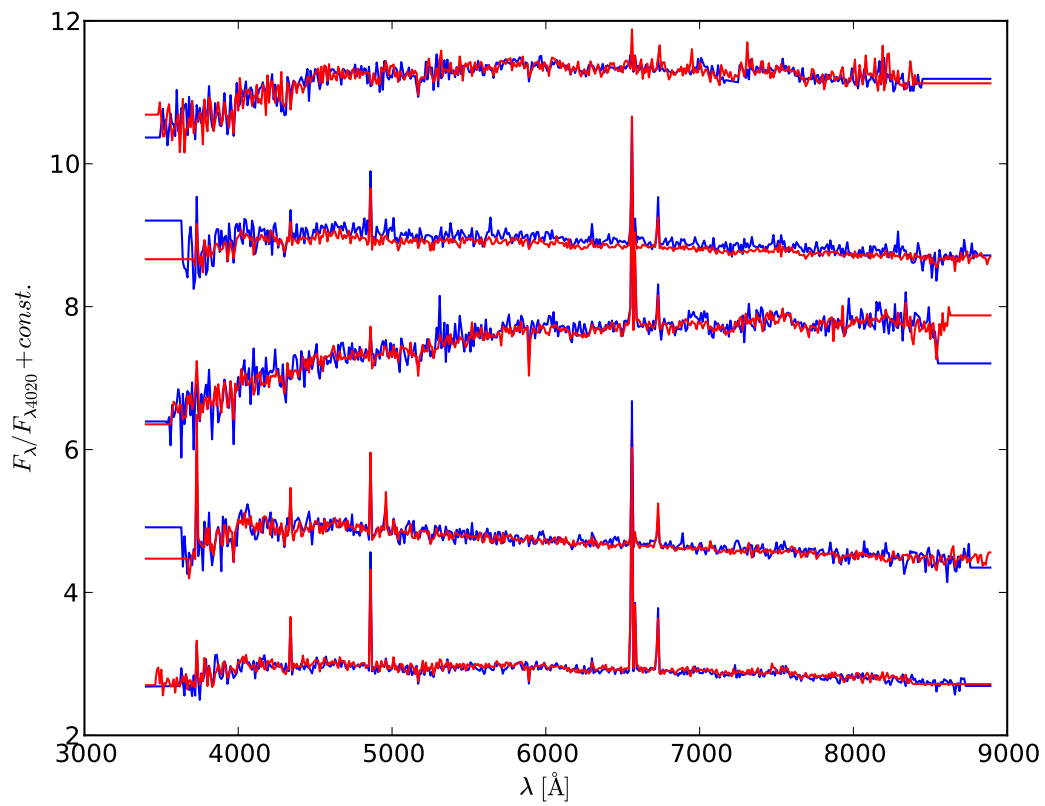


Figura 3.3: Espectros suavizados em 10 \AA dos exemplos da figura 3.2. Nesta figura pode-se evidenciar qualitativamente a precisão do ajuste, tanto no contínuo como nas linhas de emissão.

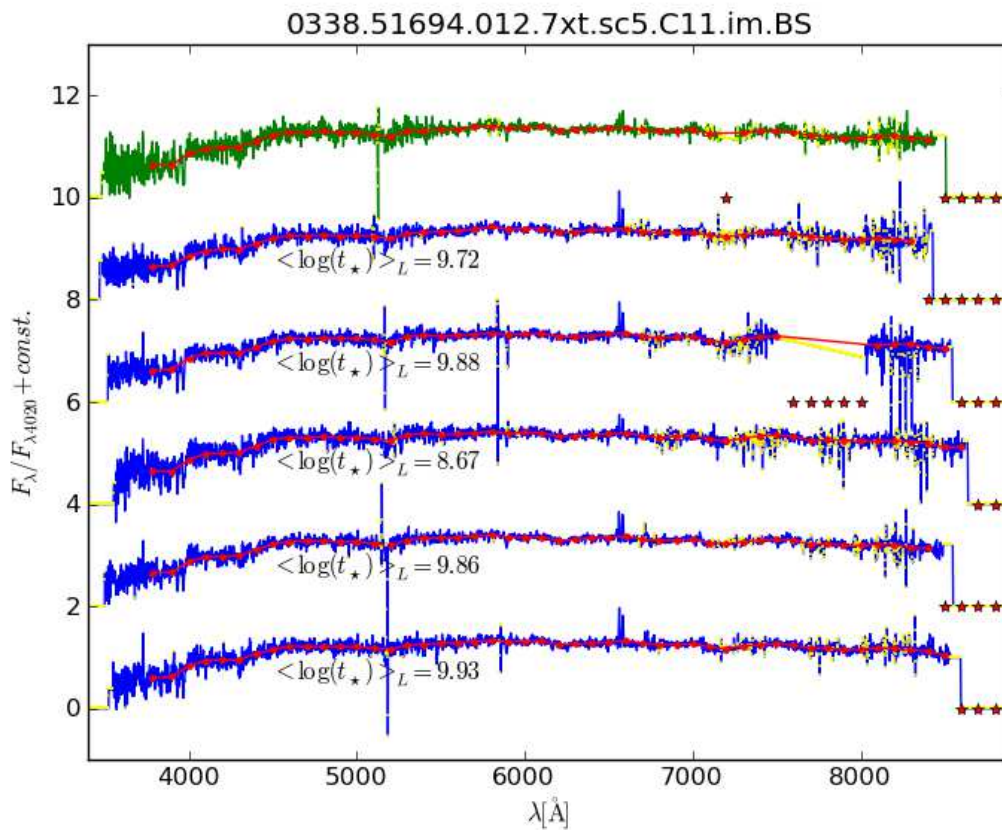


Figura 3.4: Exemplo de J-espectro observado de uma galáxia de $\langle \log t \rangle_L = 9,31$ e seus 5 melhores ajustes em χ^2 . De cima para baixo, temos em verde o objeto e, em azul, os 5 melhores ajustes a este J-espectro. Em amarelo estão marcados os pontos que não foram levados em conta por estarem mascarados pela pipeline do SDSS, devido a problemas. Em vermelho, o J-espectro correspondente. As estrelas indicam pontos em que J-espectro não foi considerado por problemas no espectro do SDSS.

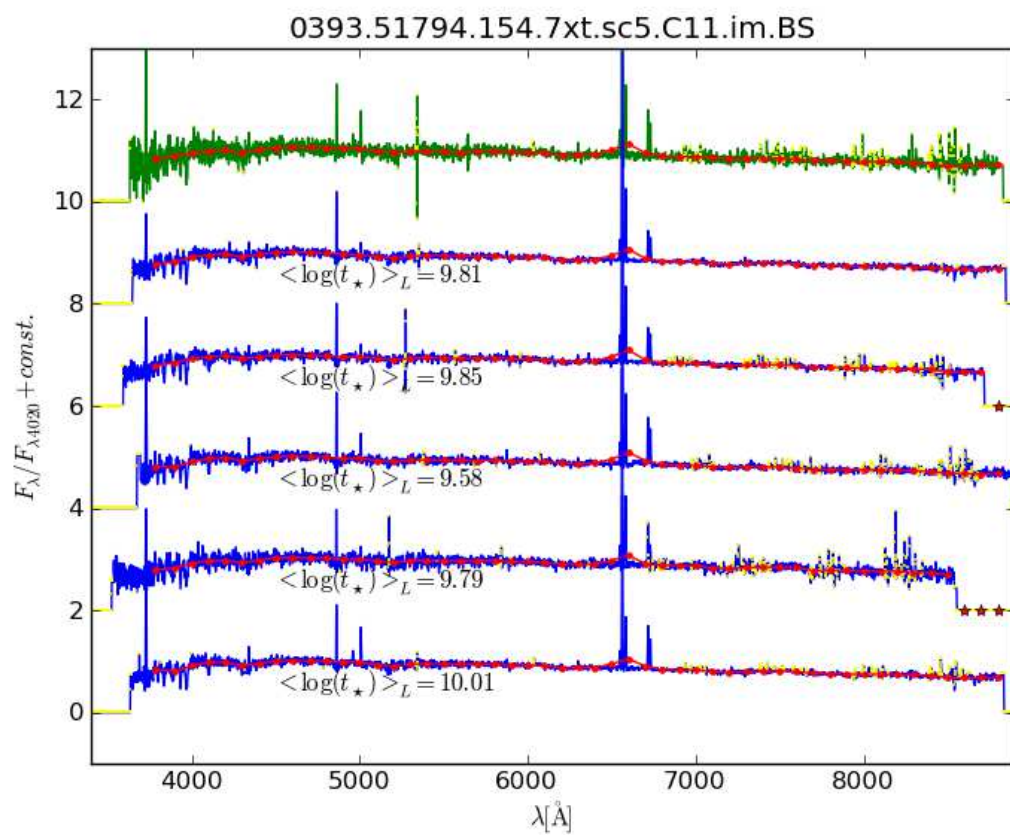


Figura 3.5: Outro exemplo de galáxia e seus 5 melhores pares encontrados, mas desta vez para um objeto mais jovem, com $\langle \log t \rangle_L = 8,50$.

que são importantes para, por exemplo, a classificação de galáxias usando diagramas de diagnóstico Baldwin, Phillips, & Terlevich 1981; Stasińska et al. 2006; Kewley et al. 2001.

A comparação realizada nesta seção foi somente uma demonstração de como podemos encontrar pares (e até conjuntos) de galáxias semelhantes, e que semelhança em J-espectros significa em termos de espectros de maior resolução (SDSS). Este resultado, mostra que a premissa deste trabalho é válida, o que nos encoraja a seguir para a próxima seção, onde desenvolvemos um formalismo que nos permite fazer inferências sobre propriedades físicas e linhas de emissão.

3.5 Formalismo Estatístico

Dados um determinado objeto e uma determinada base, gostaríamos de inferir sobre os valores das propriedades deste objeto como, por exemplo, as larguras equivalentes de linhas de emissão, idade média das populações estelares, razões de medidas de linhas de emissão e outros.

Com o valor de $\chi_{i,j}^2$ para todos os elementos da base e cada objeto, podemos calcular uma função do tipo *likelihood* ($l_{i,j} = e^{-\frac{1}{2}\chi_{i,j}^2}$), e a partir dela uma função distribuição de probabilidades $P(p)$. Como em Gallazzi et al. (2005), “A função distribuição de probabilidades para um determinado parâmetro é simplesmente dada pela distribuição naquele parâmetro das *likelihoods* de todos os modelos da biblioteca”. Em nosso caso, estamos chamando os parâmetros de **propriedades** e bibliotecas de **bases**.

Na figura 3.6, retirada de Gallazzi et al. (2005), temos um exemplo de aplicação da estatística bayesiana para a recuperação da idade média de quatro galáxias do SDSS com duas bibliotecas diferentes. Neste caso, foram comparados alguns índices espectrais ($[MgFe]'$, D_n4000 , $H_{\delta A}$, $H_{\gamma A}$, etc) como comparação aos modelos e não o espectro.

A diferença básica é que em Gallazzi et al. (2005) foram utilizados modelos de populações estelares como base e aqui estamos utilizando galáxias reais (do SDSS), o que dá ao nosso método maior robustez uma vez que estamos buscando por galáxias através da comparação com outras galáxias e não modelos. Modelos, além de não possuir importantes informações como as medidas das linhas de emissão, muitas vezes podem conter erros nas suas formulações, fazendo com que a biblioteca de modelos talvez não seja uma boa representação do universo de possibilidades.

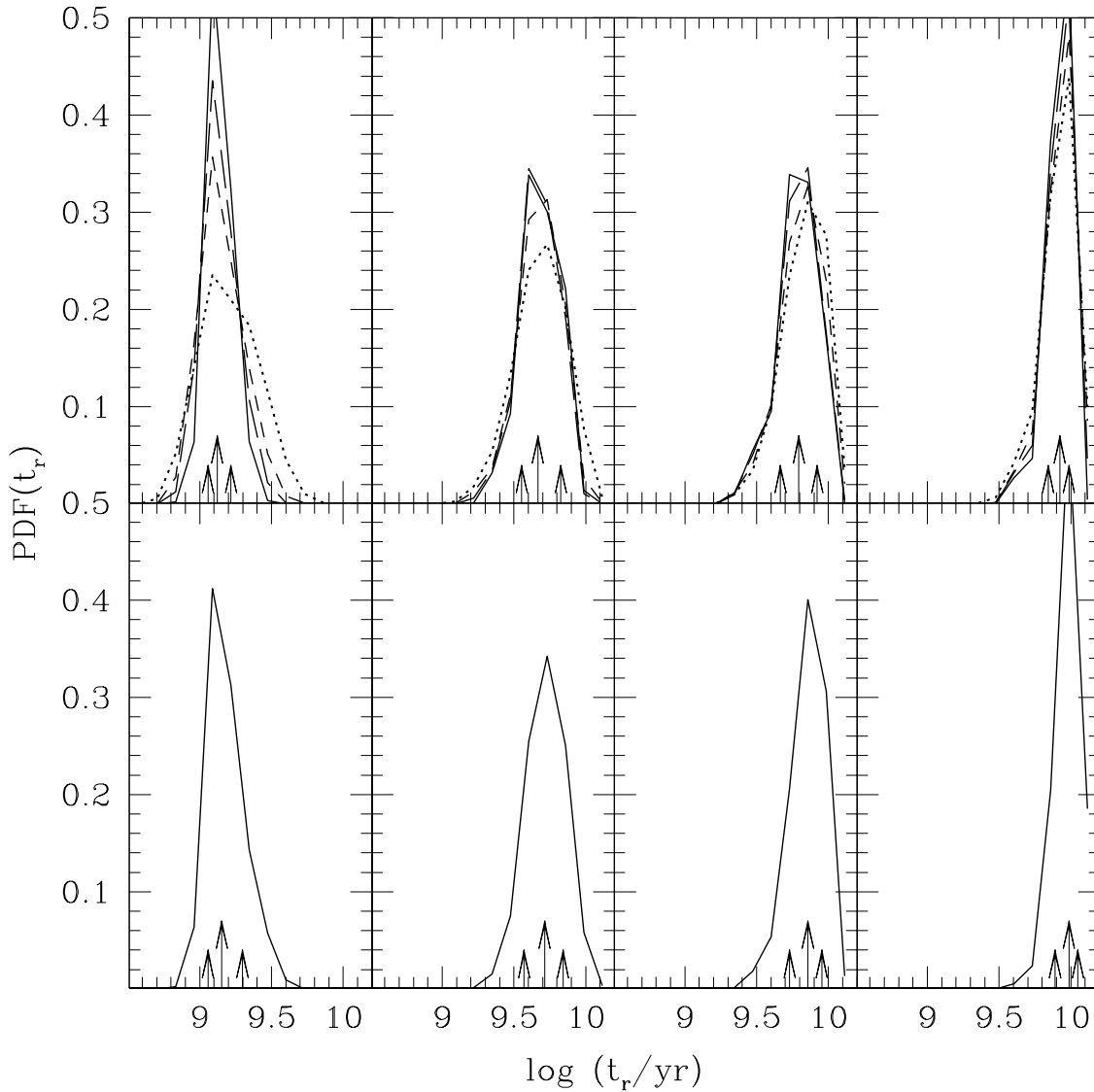


Figura 3.6: Figura retirada de Gallazzi et al. (2005). Função de distribuição de probabilidade para idade média ponderada pela luz na banda r para quatro galáxias com $S/N > 30$ e diferentes idades. As linhas sólidas do painel inferior foram obtidas com uma base que leva em consideração somente índices sensíveis à idade e, no painel superior, com uma base que leva em consideração índices que dependem também dos metais. As setas indicam o 16º percentil, a mediana e o 84º percentil, respectivamente.

De uma maneira formal (Sivia & Skilling 2006), temos que a probabilidade para uma determinada propriedade p para um dado um objeto O_i é dada por:

$$P(p|O_i) = \sum_j P(p|h_j)P(h_j|O_i) \quad (3.6)$$

onde h_j (hipótese) é um conjunto de propriedades observadas em O_i e B_j e o segundo termo desta equação é dado pelo Teorema de Bayes:

$$P(h_j|O_i) = \frac{P(O_i|h_j)P(h_j)}{P(O_i)} \quad (3.7)$$

Cada um dos termos da equação 3.7 possui um nome formal. O termo $P(h_j)$ é chamado de prior, que mede nosso estado de conhecimento (ou ignorância) sobre a hipótese antes da análise dos dados. O $P(O_i|h_j)$ é chamado de *likelihood* e mede a probabilidade dos dados dado o modelo (B , no nosso caso) e a evidência $P(O_i)$ é uma constante que desaparecerá na normalização. O termo $P(p|h_j)$ na equação 3.6 é a função distribuição de probabilidades da propriedade p para o objeto B_j da base.

Por simplicidade, assumiremos que p é medido perfeitamente (sem erros) nas galáxias da base e, portanto,

$$P(p|h_j) = \delta(p - p_j) \quad (3.8)$$

Seria mais correto (e pretendemos fazer no futuro) usar uma gaussiana centrada em p_j e com dispersão dada pela incerteza de p na galáxia j da base, dada pela incerteza observacional correspondente.

Outro termo que desaparecerá da análise é $P(h_j)$. Nosso prior será implícito, estabelecido na própria seleção de galáxias para a base (ver capítulo 4), de modo que adotaremos $P(h_j) = 1$. Em princípio, este termo poderia ser usado, por exemplo, para excluir galáxias azuis da base quando o objeto for vermelho, mas a *likelihood* deve dar conta de evitar que tal casamento tenha peso na análise. Assumindo erros gaussianos e independentes em cada filtro, a *likelihood* assume então a seguinte forma:

$$P(O_i|h_j) = \mathcal{L}_{i,j} = e^{-\frac{1}{2}\chi_{i,j}^2} \quad (3.9)$$

Podemos então reescrever a equação 3.6 como:

$$P(p|O_i) \propto \sum_j P(p|h_j) e^{-\frac{1}{2}\chi_{i,j}^2} \quad (3.10)$$

Substituímos na equação anterior o χ^2 pelo χ_{RED}^2 (equação 3.5), ficando a função distribuição de probabilidades aqui utilizada na forma:

$$P(p|O_i) \propto \sum_j P(p|h_j) e^{-\frac{1}{2}f_L \chi_{i,j,RED}^2} \quad (3.11)$$

Formalmente, o fator f_L na equação acima deveria ser igual ao inverso do número de *pixels* $N_{i,j}$ ajustado. No entanto utilizaremos f_L como mais um parâmetro técnico a ser otimizado. O valor de f_L regula a largura da função de distribuição de probabilidade aumentando ou diminuindo a quantidade de objetos da base que contribuirão significativamente para a estimativa de determinada propriedade p . Este fator será explorado posteriormente na seção 4.3.

Nas figuras 3.7 e 3.8, mostramos para uma galáxia a estimativa da idade média $\langle \log t_\star \rangle_L$ e $\log[\text{N II}]/\text{H}\alpha$ a partir de uma base qualquer usando o formalismo apresentado, onde a função distribuição de probabilidade descrita na equação 3.10 está representada pela linha sólida verde no painel superior. Cada um dos pontos do painel central representa um elemento da base. Note que o número de galáxias que contribuem efetivamente para a PDF é pequeno. Contribuem somente os poucos pontos que estão na parte inferior deste gráfico ($\log \chi^2/N \approx 0$).

Podemos observar a pequena diferença entrada versus saída na idade média, como já esperávamos, pois a cor do espectro se preserva, mas o que nos chama atenção ainda mais é conseguir recuperar a razão de linhas de emissão $\log[\text{N II}]/\text{H}\alpha$ de forma tão precisa, já que estas linhas não aparecem de forma evidente como a cor. Comparando estas diferenças (saída menos entrada) para todos os objetos de nossa amostra, podemos ter uma

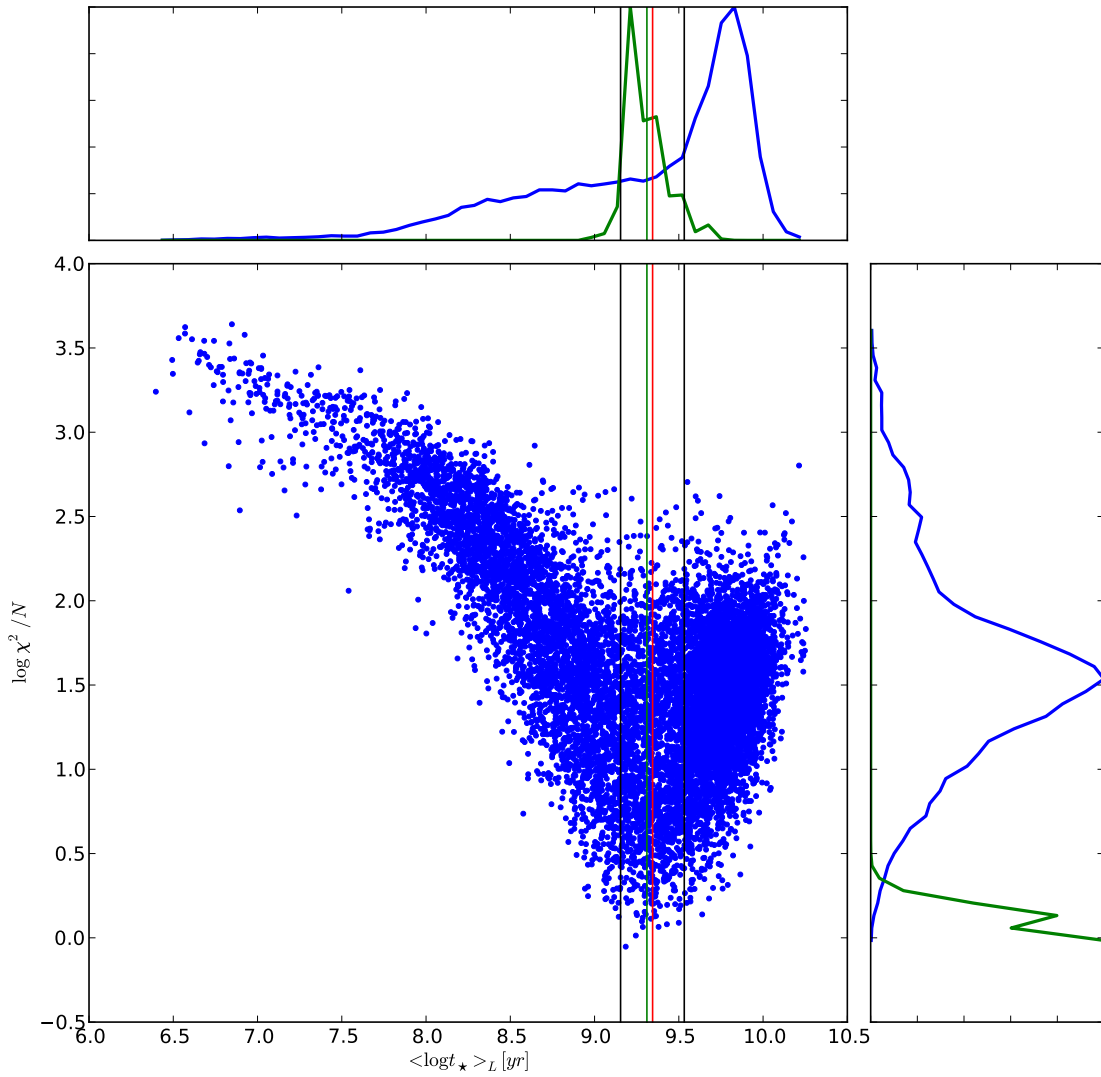


Figura 3.7: Função distribuição de probabilidade para $\langle \log t_{\star} \rangle_L$ do objeto da figura 3.4. Cada um dos pontos representa um elemento da base. No eixo x, temos a idade média ponderada pelo fluxo (extraída da síntese do STARLIGHT e no eixo y o logaritmo do χ^2 reduzido (equação 3.5). Nos histogramas lateral e superior temos em azul a distribuição para a base e em verde a mesma distribuição ponderada pela *likelihood*. As linhas verticais em preto representam os percentis de 16% e 84%, a verde é o valor da idade média que a síntese do STARLIGHT retornou para este objeto (valor de entrada) e a vermelha é a média das idades médias ponderada pela *likelihood* (valor de saída).

estimativa da precisão de nosso método. Os resultados detalhados destas comparações serão apresentados no capítulo 5.

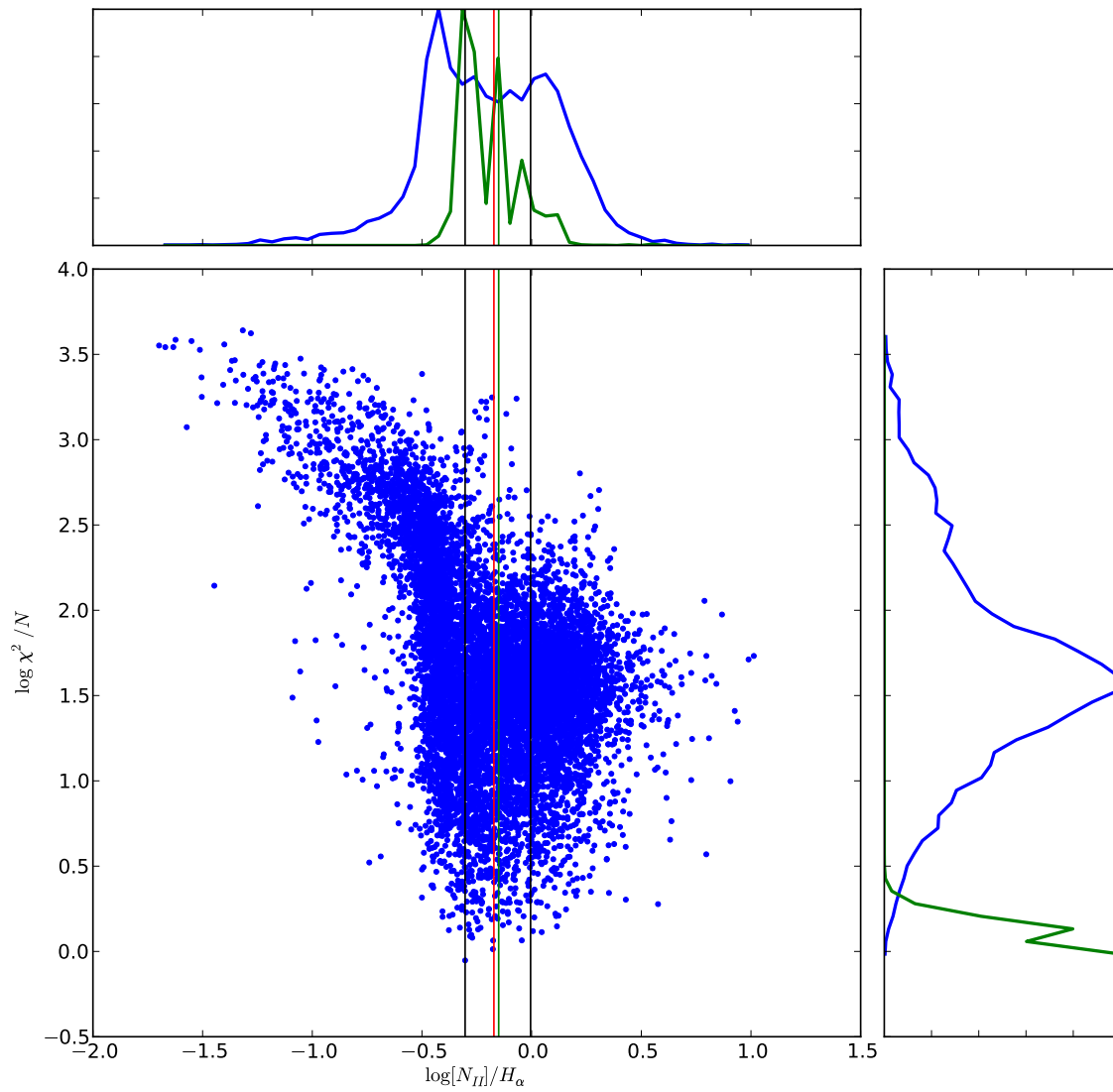


Figura 3.8: Mesma figura anterior, porém para a razão de linhas de emissão $\log[N_{II}]/H\alpha$.

Capítulo 4

Amostras, bases e detalhes técnicos

RESUMO: Neste capítulo descrevemos a amostra geral de dados do SDSS utilizada tanto para simular objetos como para a construir bases de referência na seção 4.1. Além disso, na seção 4.2 apresentamos duas bases específicas que usamos para testar o método de casamento espectral e avaliar sua precisão. Por fim, discutimos o efeito do fator f_L , e na seção 4.3 apresentamos experimentos numéricos de calibração de valores f_L otimizados para distintas finalidades. Esses desenvolvimentos completam a apresentação do método proposto neste trabalho, e abrem caminho para a apresentação dos resultados de nossas simulações, apresentados no capítulo seguinte.

4.1 Amostra Geral

Apesar de já termos apresentado vários exemplos reais no capítulos anteriores, até agora não delimitamos explicitamente a amostra de dados trabalhada nesta dissertação. Esta é a função desta seção.

Nossas amostras são totalmente oriundas do SDSS. Inicialmente construímos uma amostra geral aplicando um conjunto de critérios semelhantes aos utilizados nos últimos artigos do grupo SEAGal: Cid Fernandes et al. (2010) e Cid Fernandes et al. (2011). Especificamente, partimos da amostra de 926426 galáxias do DR7 do SDSS analisadas com o STARLIGHT, aplicando a seguinte seleção:

- As galáxias têm de ser do “Main Galaxy Sample” do SDSS;
- Caixas de 31 Å e centrada nas linhas de emissão $H\alpha$, $H\beta$, $[N II]\lambda 6584$, $[O II]\lambda 3727$

e [O III] λ 5007 não devem possuir nenhum *pixel* ruim;

- O *redshift* das galáxias deve estar entre 0,01 e 0,11

Estes critérios de seleção nos forneceram uma amostra geral de 299.253 galáxias. O último critério deve-se ao fato de querermos que os dados a faixa espectral do SDSS caiba completamente na região dos 51 filtros deste trabalho (figura 2.4). O segundo foi estabelecido a fim de, quando não houver nenhuma medida de linha de emissão, que seja porque esta é inexistente ou muito fraca e não devido a algum problema observacional.

Esta amostra reúne todos os tipos espectrais de galáxias do universo local (com formações estelares de todos os tipos, altas e baixas metalicidades, diferentes extinções, núcleos ativos de linhas de emissão estreitas, etc), com a única exceção de galáxias AGNs de linhas largas (*Seyferts* de tipo 1 e Quasares).

A partir desta seleção, dividimos as galáxias em dois grupos: um cujos espectros possuem razão sinal ruído (S/N , medida em um intervalo entre 4730 Å e 4780 Å) maior que 20 e outro com S/N menor que 20. As 113.821 galáxias com $S/N \geq 20$ formam o grupo das quais serão utilizadas como base de referência principal e as outras 185.432 com $S/N < 20$ formam o conjunto das quais serão utilizadas como objetos, cujas propriedades deverão ser estimadas a partir do casamento espectral em relação a uma subamostra das galáxias de referência.

Como todas as informações das galáxias vêm do nosso bancos de dados, temos medidas diretas e padronizadas de todas as propriedades que gostaríamos inferir de uma galáxia usando os dados do J-PAS e isto será usado para aferir a precisão do método que propomos. Em outras palavras, compararemos valores de entrada conhecidos (por exemplo, a largura equivalente de H α em um objeto) com as estimativas que obtemos a partir de nosso método.

4.2 Bases

A seleção de uma base de referência é um dos principais ingredientes do método de casamento espectral. Assim como em métodos que comparam medidas observadas com bibliotecas de modelos (Gallazzi et al. 2005), a composição de uma base que representará um universo de parâmetros assume o papel implícito de um prior (equação 3.7).

Em geral, análises bayesianas devem ser relativamente imunes à escolha do prior, no

sentido de que o resultado não deve depender criticamente deste. Neste caso é evidente que não devemos construir uma base composta somente de galáxias sem linhas de emissão ou, ao contrário, somente com galáxias com linhas de emissão fortes. Devemos ter em mente ao definir critérios de seleção para uma base que ela seja a mais genérica possível para que cubra todo os tipos de espectros de galáxias.

Tendo isto em mente, surge o problema de como implementar este “princípio filosófico” de forma objetiva. Nesta seção apresentamos duas propostas de bases que satisfazem este princípio geral de conter todos os tipos espectrais relevantes, mas são construídas de formas completamente independentes. Em ambos casos, as bases foram construídas a partir de um diagrama bidimensional, dividido em pequenas caixas nos eixos X e Y e populando cada pixel (X, Y) de acordo com uma terceira dimensão Z . Uma das bases (seção 4.2.1) usa no plano (X, Y) um análogo físico do diagrama cor-magnitude (CMD, do inglês) para galáxias, enquanto a outra (seção 4.2.2) utiliza o diagrama de diagnóstico WHAN introduzido por Cid Fernandes et al. (2011). No caso da base CMD, o terceiro eixo (Z) foi, quando possível, a extinção A_V e, no caso da base WHAN, o eixo Z escolhido, quando possível, foi a razão das linhas de emissão $\log[\text{O III}]/\text{H}\beta$. Para escolha dos elementos, ordenamos os elementos da caixa (X, Y) em ordem crescente de Z e, então, escolhemos $10\% + 1$ ao longo de Z .

4.2.1 Base de Propriedades Físicas: Diagrama “cor-magnitude”

Em astrofísica, o diagrama cor-magnitude é a base para estudos da evolução estelar. As estrelas passam sua vida ao longo deste diagrama, passando a maior parte de sua vida em uma linha bem definida chamada sequência principal, movendo-se à região das estrelas gigantes e, logo, explodindo em forma de supernovas ou convertendo-se em anãs brancas. Em galáxias (Bell et al. 2004; Baldry et al. 2004), temos uma distribuição análoga à estelar, com uma distribuição bimodal de galáxias divididas em dois grupos principais: jovens (*blue cloud*) e velhas (*red sequence*), com uma parte transitória (*green valley*). Na figura 4.1, retirada de Baldry et al. (2004), podemos ver um exemplo de um diagrama destes para 66.856 galáxias do SDSS.

Na figura 4.2, mostramos nossa versão do diagrama “CMD-físico” onde substituímos a magnitude pela massa e a cor pela idade média. Pode-se observar uma evidente correspondência entre os dois diagramas. Como massa e idade de galáxias são duas propriedades que se correlacionam muito com as outras que desejamos medir, este critério de seleção de base se justifica.

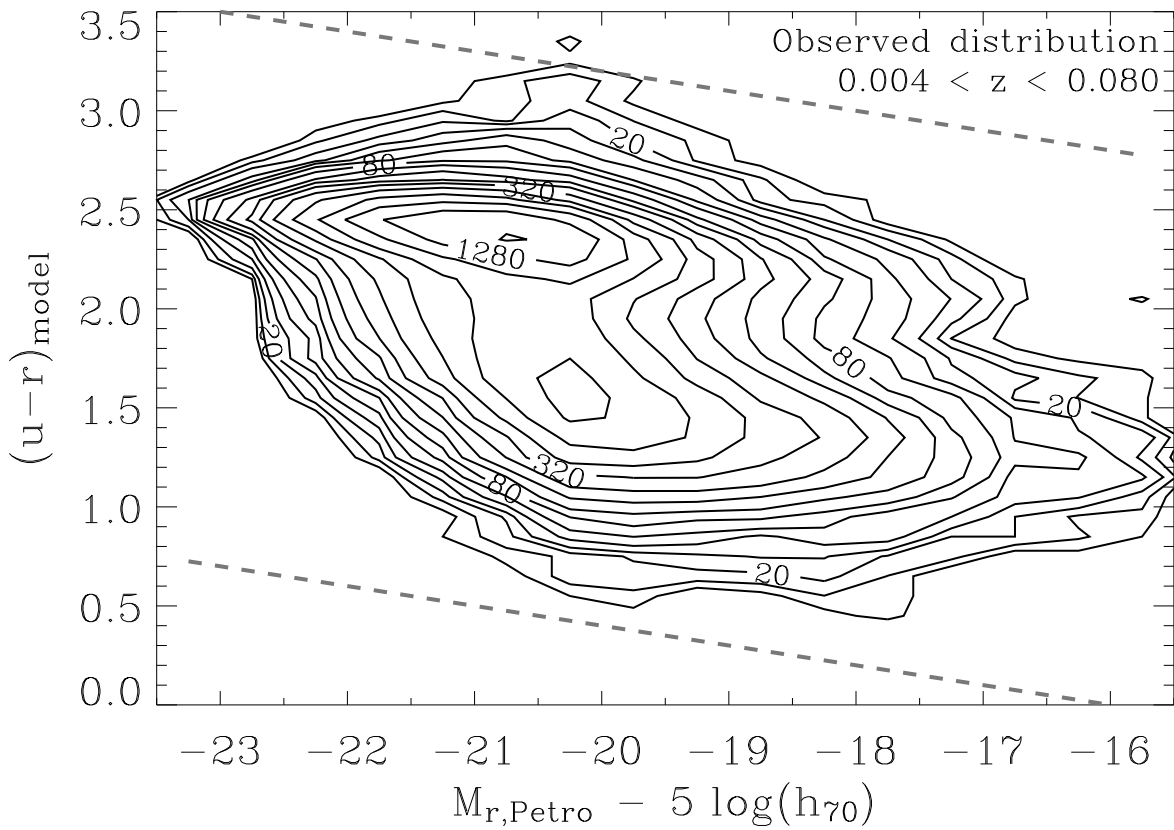


Figura 4.1: Diagrama cor/magnitude extraído de Baldry et al. (2004). Os contornos são determinados pela quantidade de galáxias contidas em caixas de 0.1 dex em cor por 0.5 em magnitude de forma logarítmica, iniciando em 10 e dobrando a cada dois contornos, com um total de 66.856 galáxias.

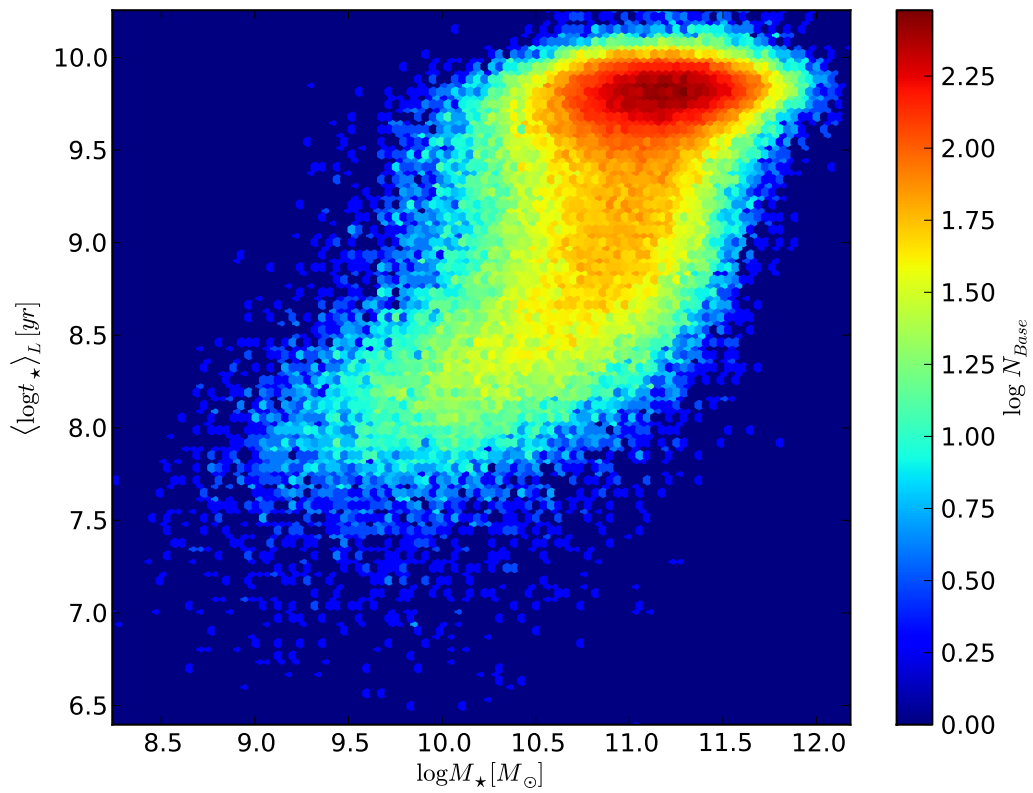


Figura 4.2: Diagrama cor-magnitude para a amostra de galáxias selecionadas como base. Esta figura foi fatiada em caixas de 0,1 dex e, para cada caixa, foram retirados 10% dos objetos variando ao longo de A_V quando possível. Nesta base, encontramos 11952 elementos.

4.2.2 Base de Linhas de Emissão: Diagrama WHAN

O diagrama cor-magnitude físico apresentado na seção anterior nos remete a propriedades estelares do espectro de uma galáxia. Porém, uma outra importante componente que faz parte de uma galáxia são as linhas de emissão que estão relacionadas ao gás no meio interestelar. Apesar de existirem correlações entre propriedades estelares e as linhas de emissão, é interessante definir uma base que levará a um prior baseado nas linhas de emissão, como já discutimos na seção 4.2.

Neste trabalho, visando uma aplicação direta da medida das linhas de emissão, temos o interesse em classificar galáxias a partir de diagramas de diagnóstico. Assim que estabelecemos nossa segunda base sobre o diagrama WHAN (Cid Fernandes et al. 2011). A escolha do WHAN ($W_{H\alpha}$ versus $[\text{N II}]/\text{H}\alpha$) como diagrama diagnóstico base se deve ao fato de que este diagrama estende o tradicional diagrama BPT (Baldwin et al. 1981) para uma amostra mais completa. O diagrama WHAN, em uma amostra qualquer, abrange um número muito maior de galáxias do que o BPT devido ao fato de não depender da medida direta das linhas de emissão $[\text{O III}]$ e, principalmente, $\text{H}\beta$ muitas vezes com baixo S/N .

O diagrama WHAN (figura 4.3), reclassifica as galáxias em cinco categorias: galáxias com formação estelar (SF), com núcleo ativo forte (sAGN), com núcleo ativo fraco (wAGN), aposentadas (RG) e passivas (PG). Como sumarizado em Cid Fernandes et al. (2011), esta nova classificação possui três virtudes principais:

- Identifica os três principais agentes ionizantes em galáxias: estrelas jovens, núcleos ativos e HOLMES (do inglês: *Hot low-mass evolved stars*).
- É baseada no mais barato e, conseqüentemente, mais inclusivo diagrama de diagnóstico, o único capaz de classificar a grande população de galáxias com linhas fracas (Cid Fernandes et al. 2010) até agora negligenciadas. De fato, ela até nos permite visualizar a região ocupada pelas galáxias passivas, característica que não é possível através de outro método tradicional (vale ressaltar que galáxias com nenhuma linha de emissão não aparecem neste diagrama).
- As fronteiras que delimitam galáxias SF de AGN e sAGN de wAGN são baseadas em transformações otimizadas de linhas divisórias baseadas em diagramas de diagnóstico tradicionais, onde as classificações funcionam bem, fazendo que em classificações com o diagrama WHAN estas sejam preservadas e, ao mesmo tempo, resolve

o problema das galáxias não classificáveis por causa de suas linhas de emissão fracas e da incapacidade dos outros métodos de distinguir núcleos ativos verdadeiros de falsos.

- Ao contrário de outros diagramas, o WHAN permite distinguir núcleos ativos fracos de galáxias aposentadas.

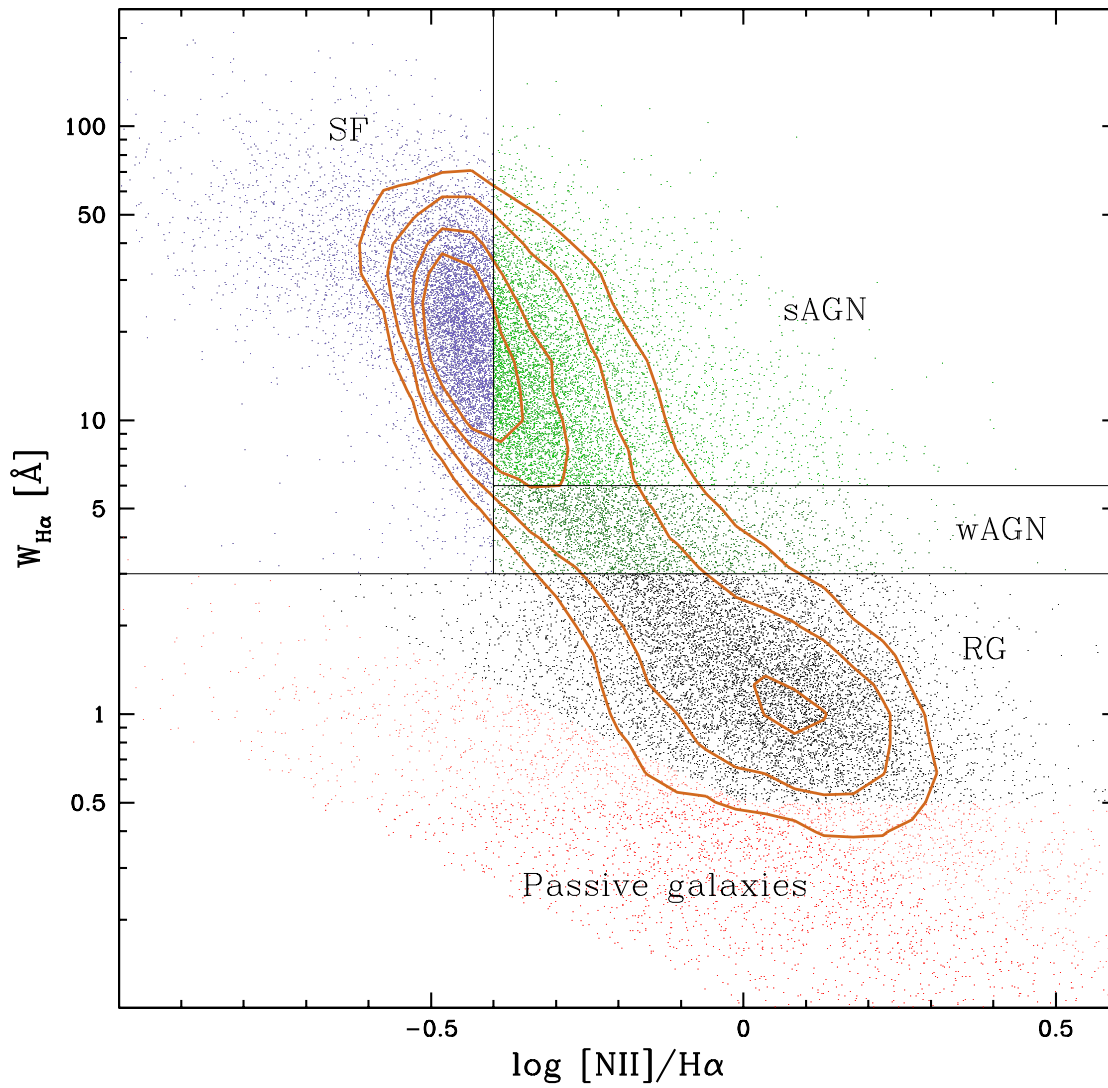


Figura 4.3: Diagrama WHAN e classificação de galáxias: galáxias com formação estelar (SF), com núcleo ativo forte (sAGN), com núcleo ativo fraco (wAGN), aposentadas (RG) e passivas (PG). Retirado de Cid Fernandes et al. (2011).

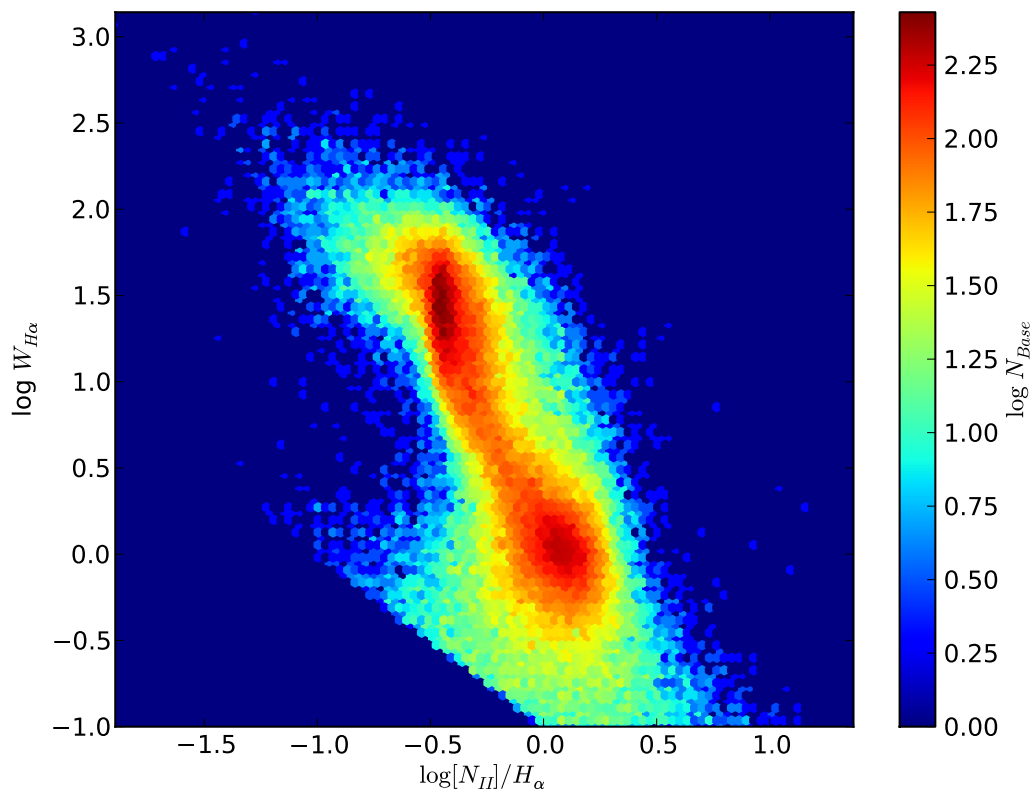


Figura 4.4: Diagrama WHAN para a amostra de galáxias selecionadas como base. Esta figura foi fatiada em caixas de 0,1 dex e, para cada caixa, foram retirados 10% dos objetos, variando ao longo de $\log[\text{O III}]/\text{H}\beta$ quando possível. Nesta base encontramos 27.536 elementos.

4.3 Otimizando o fator f_L

Antes de partir para as aplicações, temos que considerar um último ingrediente do método de casamento espectral: o fator f_L . Apesar de ser menos relevante que a escolha da base do ponto de vista filosófico, este fator tem um impacto direto na estimativa de propriedades através do método aqui proposto.

Introduzido na equação 3.10, este fator *ad-hoc* deveria, em princípio, ser da ordem do número de filtros bons no J-espectro ajustado e da base, mas na prática podemos experimentar a variação deste número para outros valores que levem a uma maior eficiência do método.

O papel prático do f_L pode ser entendido da seguinte maneira: quando f_L tende a zero, todas as galáxias tendem a contribuir igualmente nas funções distribuição de probabilidade da equação 3.11. Isto representa uma função *likelihood* que diz muito pouco sobre a comparação e o posterior será quase igual ao prior. No outro extremo, com f_L tendendo ao infinito, apenas a galáxia com o menor χ^2 contribuirá significativamente para o posterior. Neste limite, o resultado do método será igual ao do pareamento discutido na seção 3.4. Valores intermediários de f_L definirão se mais ou menos galáxias contribuirão efetivamente na conta dos posteriores.

Para estabelecer o valor ideal para f_L , adotaremos uma abordagem heurística, determinando o f_L como simplesmente o que nos remete aos melhores resultados. Por “melhor resultado” queremos dizer aquele que minimize (em um sentido estatístico) a diferença entre os valores de entrada e os de saída (os estimados bayesianamente) para uma dada propriedade dentre as que desejamos recuperar. Isto implica na possibilidade de termos um f_L ótimo para cada propriedade, ou seja, $f_L = f_L(p)$, onde p é uma propriedade qualquer.

Para determinar o f_L ideal, foram calculadas as dispersões das diferenças entre os parâmetros de entrada (provenientes do STARLIGHT) e a média de determinada propriedade ponderada pela função distribuição de probabilidades (descrita mais formalmente na próxima seção, equação 4.2). Assim, o fator f_L a ser escolhido para cada propriedade p será o que minimiza a relação abaixo:

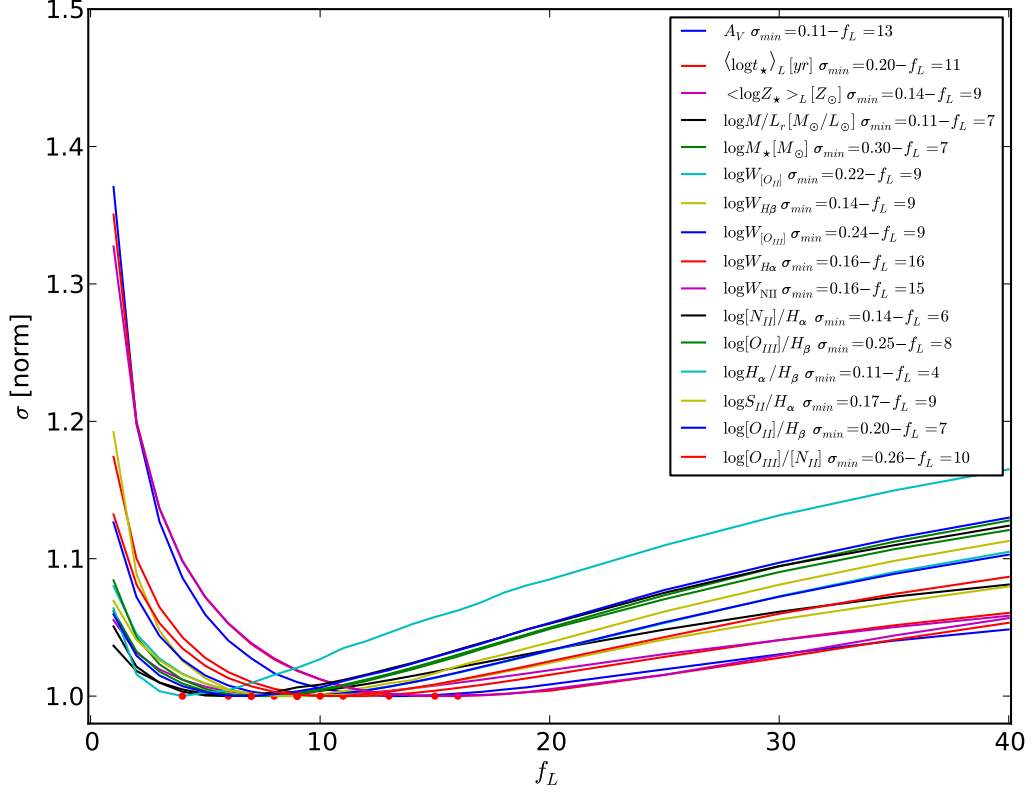


Figura 4.5: Dispersão da diferença entre entrada e saída (média, equação 4.2) para diferentes fatores f_L e diferentes propriedades para a base CMD. Para este gráfico, foram calculadas estatísticas para $f_L = 1$ a $f_L = 20$ com $\Delta f_L = 1$ e para $f_L = 25, 30, 35$ e 40 . Na legenda, estão representados os valores mínimos para a dispersão de cada valor e o f_L que representa estes mínimos.

$$\sigma(p_{f_L}) = \sqrt{\frac{1}{N} \sum_i (p_{f_L, \text{saida}, i} - p_{f_L, \text{entrada}, i})^2} \quad (4.1)$$

Isto foi realizado para ambas bases, e os resultados são mostrados nas figuras 4.5 e 4.6. Para ambas as bases, encontramos valores de f_L da ordem de 10 para as diferentes propriedades físicas e linhas de emissão. Os valores aqui encontrados serão utilizados nas simulações apresentadas no capítulo 5.

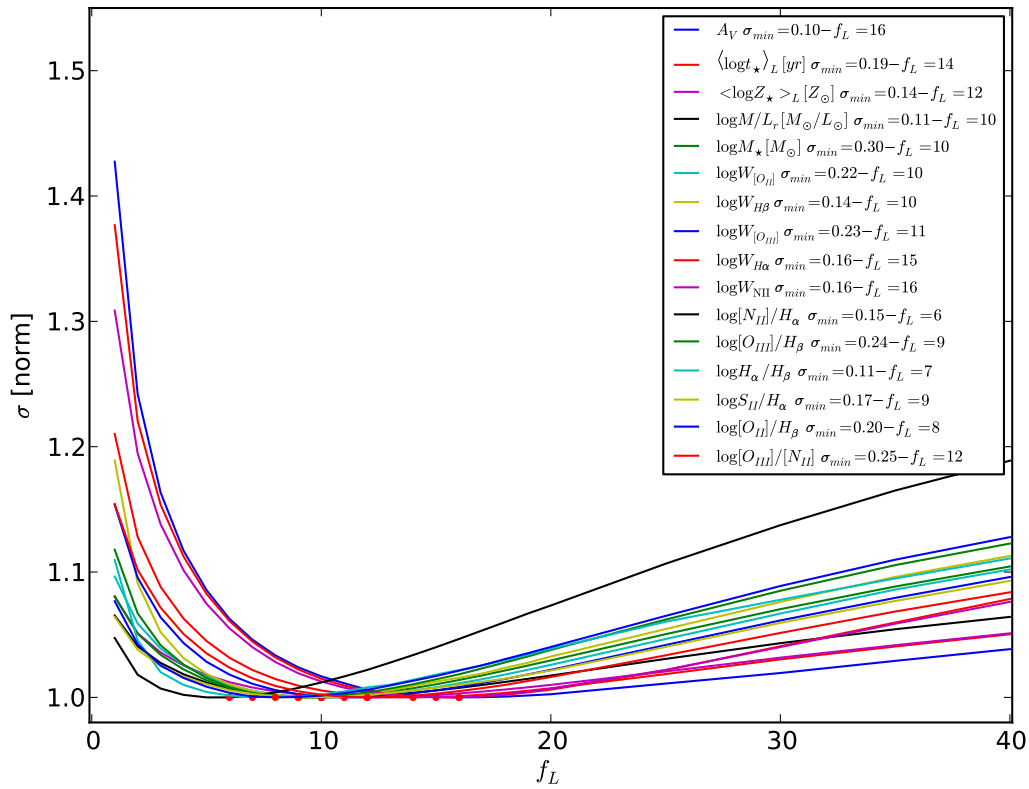


Figura 4.6: Mesmo gráfico anterior, mas para a base “WHAN”.

4.4 Estimadores: Resumos da Função Distribuição de Probabilidades

Mesmo com toda a elegância que análises probabilísticas proporcionam, no sentido que a função distribuição de probabilidade é formal e filosoficamente a maneira mais correta de expressar os resultados, para realizarmos comparações concretas entrada versus saída necessitamos um número que quantifique determinada propriedade. Na prática, com uma função distribuição de probabilidades não podemos responder perguntas do tipo: “Qual a largura equivalente de $H\alpha$ do objeto número tal do J-PAS?”. Assim se torna necessário resumir esta função em um número que expresse a quantidade que desejamos medir mais um outro que represente o erro (ou a confiança) nesta medida.

Há várias maneiras de se fazer isto. Neste trabalho, utilizaremos a média e o desvio padrão como representações destes valores. Assim, nas comparações entrada versus saída no capítulo seguinte, os valores que adotaremos como saída de nosso método serão, na realidade, a média da propriedade dada a função distribuição de probabilidades correspondente à equação 4.2. E, quando falamos nos erros de saída, estamos nos referindo ao desvio padrão desta função dado pela equação 4.3:

$$\bar{p}_i = \frac{\sum_j p_j \mathcal{L}_{i,j}}{\sum_j \mathcal{L}_{i,j}} = p_{saida,i} \quad (4.2)$$

$$\sigma(p_i) = \frac{\sqrt{\sum_j (p_j - \bar{p}_i)^2 * \mathcal{L}_{i,j}}}{\sum_j \mathcal{L}_{i,j}} \quad (4.3)$$

lembrando que $\mathcal{L}_{i,j} = e^{-\frac{1}{2}f_L \chi_{i,j}^2}$, como definido na equação 3.9.

Por exemplo, para o objeto das figuras 3.7 e 3.8, estimamos $\langle \log t_\star \rangle_L = 9,34 \pm 0,18$ e $\log[N_{II}]/H\alpha = -0,15 \pm 0,18$. Além de estimar valores para propriedades desejadas, nosso método proporciona também uma estimativa de erros através da PDF, o que é uma virtude do método. No entanto, este fato não será explorado neste trabalho devido à aproximação feita na equação 3.8.

Capítulo 5

Resultados

RESUMO: *Nos capítulos anteriores apresentamos os princípios gerais do nosso método, seus ingredientes, o formalismo e os detalhes técnicos. Neste capítulo, aplicamos a metodologia desenvolvida com o objetivo de testar e quantificar quão precisamente podemos recuperar uma série de propriedades de galáxias.*

5.1 Propriedades de interesse

O foco deste trabalho é testar se o casamento espectral é capaz de nos fornecer informações úteis sobre os nossos objetos alvo. Com a metodologia apresentada nos capítulos anteriores, neste capítulo verificaremos se estas propriedades podem ser recuperadas e com qual precisão. A precisão será aferida comparando-se a saída de nosso método com os valores conhecidos, provenientes da análise com o STARLIGHT, filosofia típica utilizada em qualquer simulação: comparar a entrada com a saída. Vale lembrar que o p_{saida} utilizado neste capítulo é definido média ponderada pela PDF, dada pela equação 4.2.

Estas comparações se centrarão em um conjunto de propriedades físicas e observacionais que são de interesse e que acreditamos poder recuperar. Dentre as muitas propriedades que dispomos em nosso banco de dados, investigaremos as listadas na tabela 5.1.

Tabela 5.1: Propriedades analisadas, suas unidades e descrições.

Propriedade	Descrição	Unidade
A_V	Extinção estelar	mag
$\langle \log t_\star \rangle_L$	Idade estelar média ponderada pela luz	anos
$\langle \log Z_\star \rangle_L$	Metalicidade estelar média ponderada pela luz	Z_\odot
$\log M/L_r$	Razão massa-luminosidade na banda r do SDSS	M_\odot/L_\odot
$\log M_\star$	Massa total	M_\odot
$\log W_{[OII]}$	Largura equivalente de $[OII]$	\AA
$\log W_{H\beta}$	Largura equivalente de H_β	\AA
$\log W_{[OIII]}$	Largura equivalente de $[OIII]$	\AA
$\log W_{H\alpha}$	Largura equivalente de H_α	\AA
$\log W_{[NII]}$	Largura equivalente de $[NII]$	\AA
$\log[NII]/H_\alpha$	Razão de fluxo de linhas de emissão $[NII]$ por H_α	-
$\log[OIII]/H_\beta$	Razão de fluxo de linhas de emissão $[OIII]$ por H_β	-
$\log H_\alpha/H_\beta$	Razão de fluxo de linhas de emissão H_α por H_β	-
$\log[SII]/H_\alpha$	Razão de fluxo de linhas de emissão SII por H_α	-
$\log[OII]/H_\beta$	Razão de fluxo de linhas de emissão $[OII]$ por H_β	-
$\log[OIII]/[NII]$	Razão de fluxo de linhas de emissão $[OIII]$ por $[NII]$	-

Esta lista contempla um conjunto de propriedades que diagnosticam tanto as propriedades estelares dos objetos (como idade média das populações, metalicidade e poeira) como as propriedades do gás (através das medidas das linhas de emissão e razões de seus fluxos). Não é demais ressaltar novamente a importância da medida de linhas de emissão pelo nosso método, uma vez que propriedades como extinção, idade média podem ser medidas utilizando-se métodos tradicionais de comparação com bibliotecas de modelos (Gallazzi et al. 2005; Brinchmann et al. 2004), mas linhas de emissão não podem. Preferimos aqui converter todas as propriedades para sua forma logarítmica (exceto A_V que já é de natureza logarítmica) para que as dispersões nas análises de entrada versus saída fossem sempre relativas.

Note que a massa estelar total (M_*) incluída nesta tabela é calculada a partir de uma extrapolação para a quantidade de luz que está fora dos 3" de abertura da fenda do espectrógrafo. Além disso, por ser uma quantidade extensiva, não esperamos reproduzi-la (duas galáxias podem ter espectros idênticos, mas massas distintas). Mesmo assim, a incluímos na análise a título de curiosidade. Para calcular massas de galáxias, utilizaremos a razão massa-luminosidade, que independe destes fatores.

5.2 Resultados: Propriedades Físicas

Na figura 5.1, mostramos histogramas da diferença saída menos entrada para as propriedades físicas listadas na tabela 5.1. Nesta figura, a base de referência utilizada foi a CMD (seção 4.2.1) em uma amostra de cem mil objetos. Na tabela 5.2, resumimos a precisão destas medidas em termos da média, mediana e desvio padrão da distribuição de pontos saída menos entrada. Tanto para a extinção quanto para a idade média ponderada pela luz temos uma distribuição relativamente normal das diferenças saída versus entrada, porém notamos que tanto para a metalicidade média ponderada pela luz quanto para a razão massa-luminosidade a saída em geral é ligeiramente menor que a entrada, mas o efeito é pequeno.

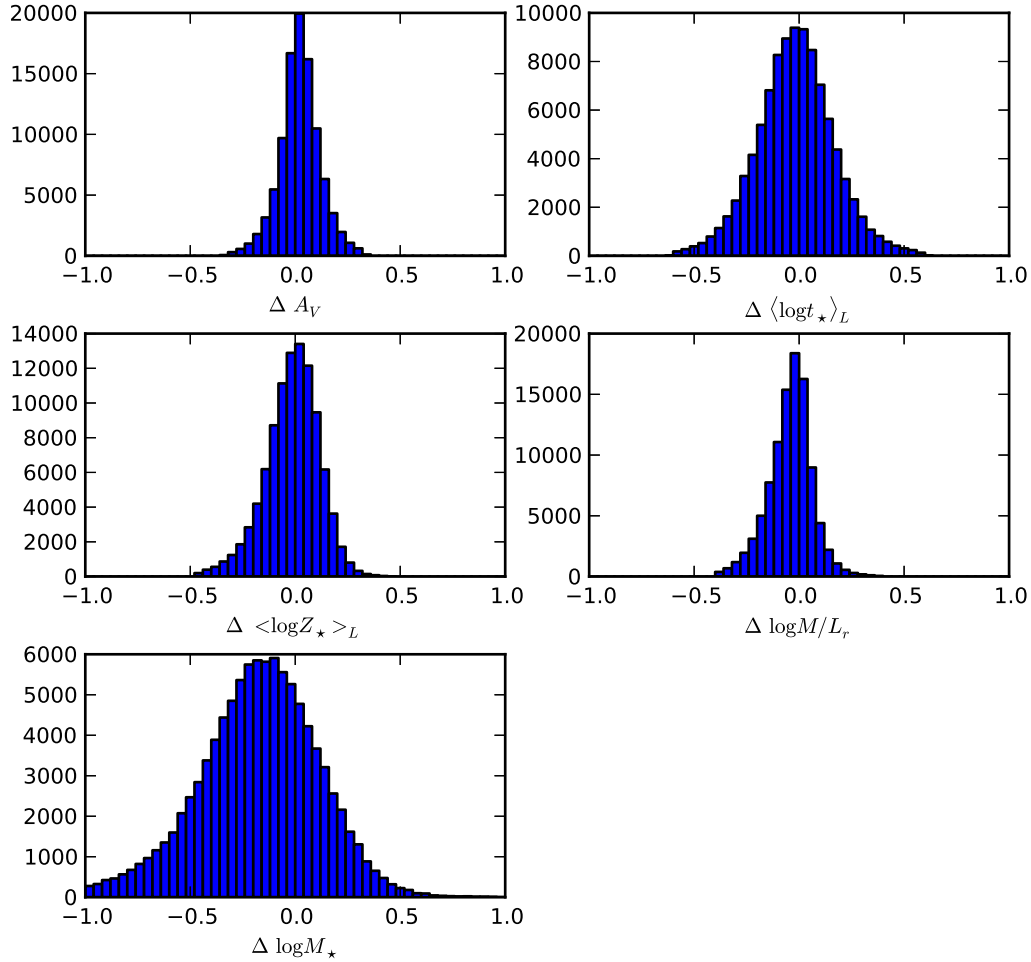


Figura 5.1: Histogramas de saída menos entrada para as propriedades físicas de interesse e base CMD.

A tabelas 5.2 e 5.3 resumem os detalhes de nossas análises. Temos para cada propriedade a diferença saída menos entrada (Δp) típica, mediana ($\tilde{\Delta p}$), dispersão ($\sigma(\Delta p)$), mínimos e máximos dos elementos da base. Nestas tabelas, a informação mais relevante está na quinta coluna que nos reporta qual a precisão média com que recuperamos as propriedades físicas. Vê-se aí que esta precisão varia entre 0,1 e 0,2 dex, o que é evidentemente bom. Curiosamente, esta precisão não varia com a base. Tanto a base CMD quanto a WHAN nos fornecem resultados muito parecidos. Note também que, como esperado, os resultados são bem piores para a massa, mas isso não é importante já que o estimador real de massa que utilizaremos será obtido através da razão massa-

luminosidade, que recuperamos com precisão de 0,11 dex.

Tabela 5.2: Melhor f_L , média, mediana, desvio padrão para $\Delta p = p_{saida} - p_{entrada}$, assim como mínimo e máximo para as propriedades físicas analisadas com a base CMD

Propriedade	f_L	$\overline{\Delta p}$	$\tilde{\Delta p}$	$\sigma(\Delta p)$	Mínimo	Máximo
A_V	13	0.023	0.022	0.106	-0.804	2.927
$\langle \log t_\star \rangle_L$	11	-0.018	-0.017	0.199	6.000	10.255
$\langle \log Z_\star \rangle_L$	9	-0.021	-0.006	0.144	-2.292	0.398
$\log M/L_r$	7	-0.045	-0.034	0.114	-1.716	0.682
$\log M_\star$	7	-0.184	-0.165	0.301	7.146	12.125

Tabela 5.3: Tabela igual a 5.2, mas para base WHAN

Propriedade	f_L	$\overline{\Delta p}$	$\tilde{\Delta p}$	$\sigma(\Delta p)$	Mínimo	Máximo
A_V	16	0.023	0.020	0.101	-0.804	2.927
$\langle \log t_\star \rangle_L$	14	-0.013	-0.012	0.192	6.000	10.255
$\langle \log Z_\star \rangle_L$	12	-0.020	-0.006	0.141	-2.292	0.398
$\log M/L_r$	10	-0.040	-0.029	0.110	-1.716	0.807
$\log M_\star$	10	-0.182	-0.164	0.296	7.146	12.125

5.3 Resultados: Linhas de Emissão

As figuras 5.2 e 5.3 são análogas à figura 5.1, porém para linhas de emissão e razões de linhas de emissão listadas na tabela 5.1. Temos resumidas as precisões nas medidas de linha de emissão e suas razões na tabela 5.4 para a base CMD e na 5.5 para a base WHAN.

Como fica evidente, os resultados são excelentes. Por exemplo podemos recuperar $\log H\alpha / H\beta$ com precisão de 0,1 dex, enquanto $\log[\text{N II}]/\text{H}\alpha$ e $W_{H\alpha}$ são recuperadas

com precisão apenas levemente inferior (0,15 dex). Este grau de precisão permite medir extinção nebular, classificação entre SF e AGN, taxa de formação estelar específica e outros. Isto partindo somente da medida direta nos dados fotométricos seria impossível. Outras linhas são recuperadas com precisões da mesma ordem.

Ao contrário do que pensávamos no início, a mudança de base de uma fundamentada em propriedades físicas para uma em linhas de emissão influenciou muito pouco quanto à precisão das medidas das linhas de emissão. Isto pode estar relacionado ao fato de haver uma grande interseção entre as duas bases ou ainda que estas estão muito sobredimensionadas. Como é comentado adiante na seção 6.2, um esforço no sentido da redução da base deverá ser aplicado antes e durante a inclusão de novos ingredientes computacionalmente mais intensivos como a estimativa do *redshift*.

Tabela 5.4: Tabela igual a 5.2, agora para linhas de emissão

Propriedade	f_L	$\overline{\Delta p}$	$\tilde{\Delta p}$	$\sigma(\Delta p)$	Mínimo	Máximo
$\log W_{[OII]}$	9	0.051	0.065	0.223	-0.971	2.232
$\log W_{H\beta}$	9	0.024	0.020	0.145	-1.000	2.119
$\log W_{[OIII]}$	9	0.046	0.048	0.245	-1.000	2.924
$\log W_{H\alpha}$	16	0.010	0.008	0.160	-1.000	2.881
$\log W_{[NII]}$	15	-0.028	-0.024	0.159	-1.000	2.282
$\log[NII]/H\alpha$	6	-0.045	-0.042	0.141	-2.099	1.140
$\log[OIII]/H\beta$	8	0.026	0.027	0.250	-1.603	1.350
$\log H\alpha/H\beta$	4	-0.011	-0.013	0.107	-0.990	2.268
$\log[SII]/H\alpha$	9	-0.006	0.019	0.172	-2.381	1.199
$\log[OII]/H\beta$	7	0.036	0.049	0.202	-1.150	1.833
$\log[OIII]/[NII]$	10	0.075	0.063	0.265	-1.944	2.098

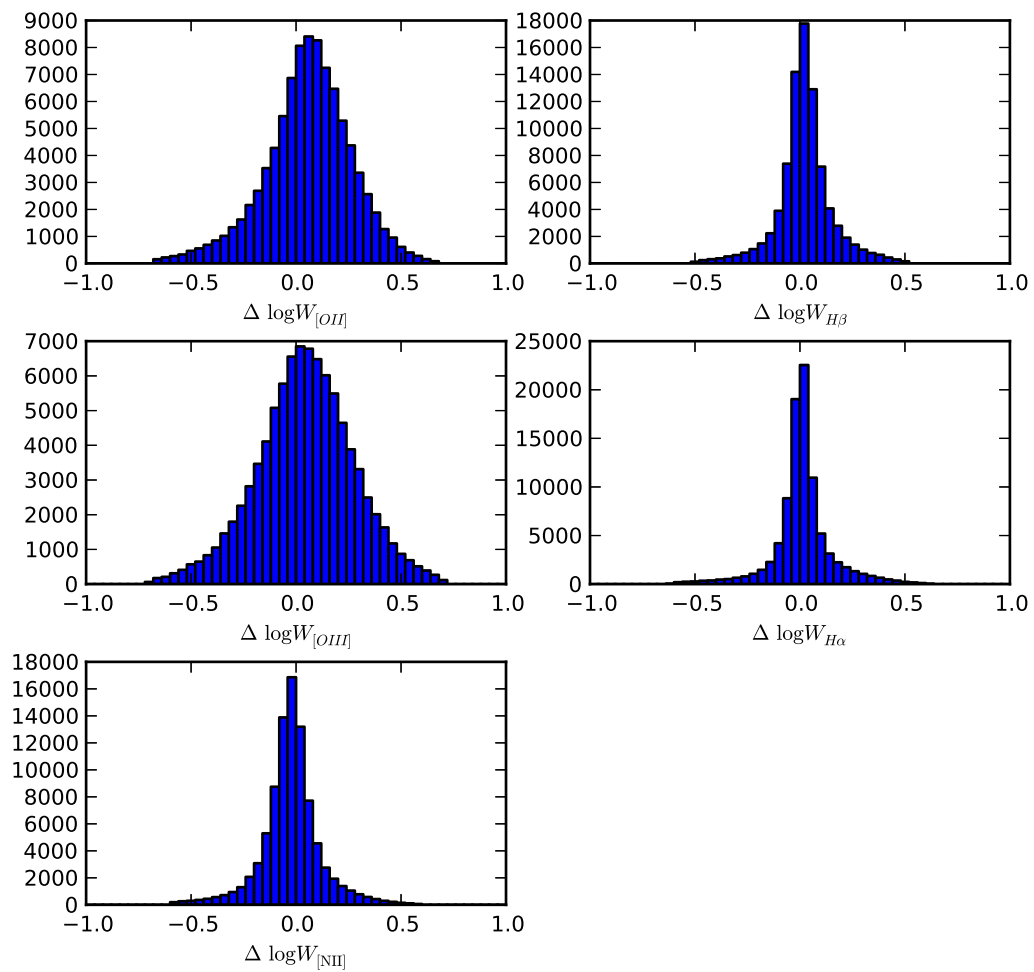


Figura 5.2: Histogramas saída menos entrada para larguras equivalentes de linhas de emissão e base CMD.

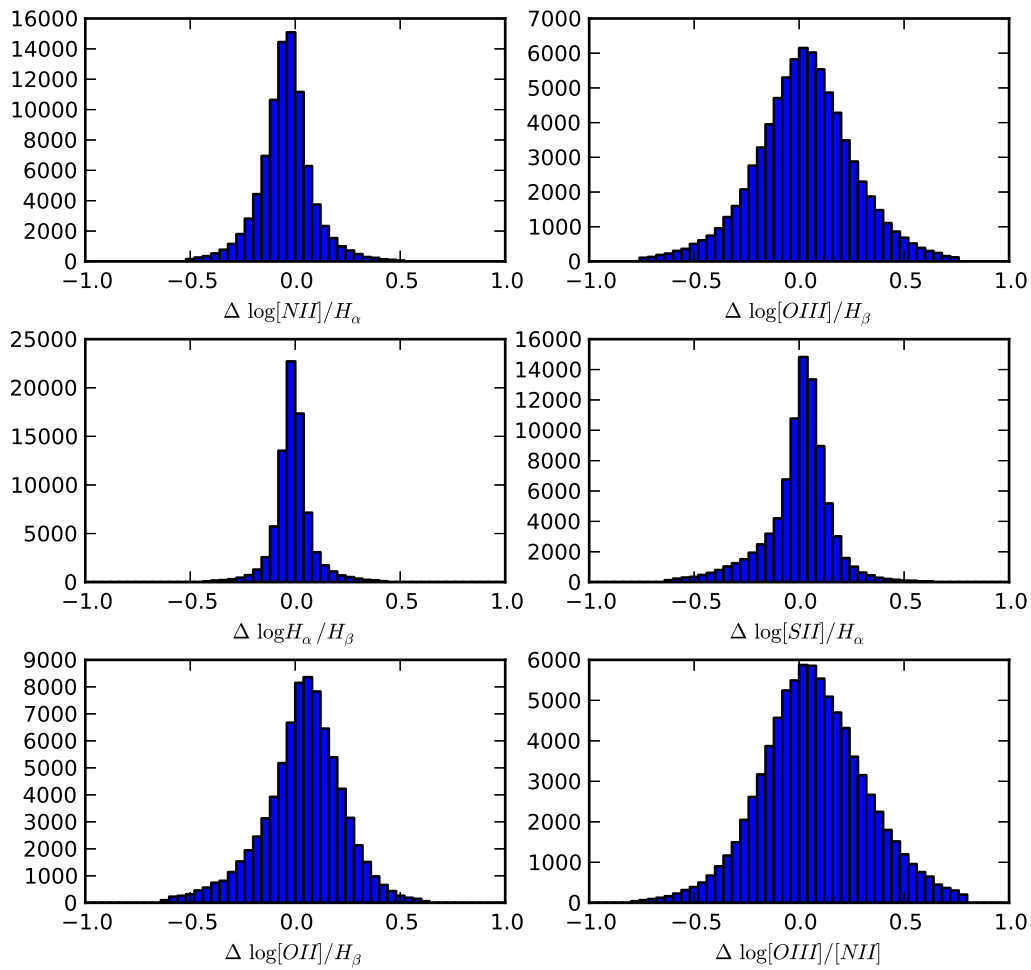


Figura 5.3: Histogramas saída menos entrada para as razões linhas de emissão e base CMD.

Tabela 5.5: Mesma tabela anterior (5.4), mas para base WHAN

Propriedade	f_L	$\overline{\Delta p}$	$\tilde{\Delta p}$	$\sigma(\Delta p)$	Mínimo	Máximo
$\log W_{[OII]}$	10	0.040	0.052	0.218	-0.971	2.232
$\log W_{H\beta}$	10	0.030	0.027	0.143	-1.000	2.119
$\log W_{[OIII]}$	11	0.029	0.034	0.232	-1.000	2.924
$\log W_{H\alpha}$	15	0.022	0.015	0.157	-1.000	2.881
$\log W_{[NII]}$	16	-0.010	-0.013	0.156	-1.000	2.282
$\log[NII]/H\alpha$	6	-0.044	-0.041	0.146	-2.099	1.140
$\log[OIII]/H\beta$	9	-0.000	0.004	0.238	-1.603	1.344
$\log H\alpha/H\beta$	7	-0.010	-0.012	0.114	-1.089	2.268
$\log[SII]/H\alpha$	9	-0.016	0.011	0.174	-2.381	1.227
$\log[OII]/H\beta$	8	0.016	0.028	0.198	-1.158	1.833
$\log[OIII]/[NII]$	12	0.044	0.037	0.252	-1.944	2.098

5.4 Resumo

Estes experimentos ilustram a pontencialidade do método aqui proposto, e cumprem nosso objetivo central de aferir a precisão com que ele nos permite inferir propriedades de galáxias a partir de J-espectros:

- Podemos medir as propriedades físicas com precisão variando entre 0,1 e 0,2 dex e linhas de emissão e razões de linhas de emissão com precisão variando entre 0,1 e 0,3 dex.
- Ao contrário que esperávamos intuitivamente, obtemos uma precisão maior na medida da metalicidade que nas idades. Contudo, este efeito se reduz quando notamos que idades cobrem 4 ordens de magnitude, enquanto metalicidade apenas 2,5.
- Nossas escolhas de base não foram relevantes para a medida tanto de propriedades físicas quanto linhas de emissão. Isto deverá ser analisado futuramente a fim de realizar uma compressão no tamanho da base, diminuindo o esforço computacional e permitindo a inclusão de outros parâmetros, como o *redshift*.

Capítulo 6

Conclusões e Perspectivas

RESUMO: Neste capítulo apresentamos um resumo do trabalho desenvolvido, seus resultados principais e enumeramos os próximos passos para o seguimento do trabalho.

6.1 Resumo do trabalho e seus principais resultados

O SDSS foi até hoje o maior levantamento de dados espectroscópicos do qual extraímos uma grande quantidade de informação sobre as propriedades das galáxias. No entanto, levantamentos espectroscópicos são caros e, pelo menos para a próxima década, não é realista se esperar a realização de outro projeto de mesmo porte para galáxias mais fracas (distantes) nem para espectroscopia espacialmente resolvida de galáxias. Apesar de haver planos nestes sentidos, levantamentos fotométricos serão em um futuro próximo muito mais abrangentes e abundantes. Boa parte da motivação para a realização destes levantamentos é cosmológica e têm seus esforços direcionados à medida precisa de *redshifts* de milhões de galáxias para estimar Oscilações Acústicas Bariônicas e propriedades da matéria escura (Benítez et al. 2009). Neste trabalho, propomos uma forma de aproveitar esta massa gigantesca de dados em prol de outra causa: medir propriedades das populações estelares e linhas de emissão nas galáxias.

Desenvolvemos e testamos uma técnica de casamento espectral para extrair este tipo de informação de dados de fotometria para 56 filtros estreitos iguais aos planejados para o projeto JPAS. O método consiste em comparar os fluxos nestes filtros (chamados aqui de J-espectros) observados com aqueles medidos em amostras de referência obtidas com o SDSS e para as quais dispomos de um grande número de propriedades muito

bem medidas em espectros com resolução muito maior do que será obtida com o JPAS. A premissa principal é que, se os J-espectros de duas galáxias são parecidos, os seus espectros em maior resolução também serão e, conseqüentemente, suas propriedades.

Assim, no capítulo 2 apresentamos os dois levantamentos (SDSS e JPAS) e, em seguida, apresentamos como foram convertidos os espectros do SDSS para os do JPAS. No capítulo 3 introduzimos o método de casamento espectral e apresentamos o formalismo estatístico adotado neste, incluindo alguns exemplos.

Um ingrediente essencial neste método é a base de comparação. Formalmente, ela entra como um prior implícito na análise. Em princípio, diferentes bases podem ser construídas, dependendo do que se quer inferir a partir dos J-espectros. No capítulo 4 apresentamos duas bases com as quais trabalhamos: uma construída a partir do diagrama massa-idade (análogo ao amplamente usado diagrama cor-magnitude para galáxias) e outra baseada no diagrama WHAN, que usa as linhas de emissão de $H\alpha$ e $[N II]$ para diagnosticar que tipo de agente ionizante domina a produção de emissão nebular em galáxias. Este capítulo também define quais amostras do SDSS foram utilizadas como objeto e base em nossas simulações do método e, por fim, a calibração do fator f_L que regula o peso dado a cada posterior.

A principal questão a ser respondida neste trabalho é com que precisão podemos recuperar uma dada propriedade de um objeto. Os resultados, apresentados no capítulo 5, podem ser resumidos desta forma:

- Propriedades Físicas: Medimos propriedades físicas como A_V , $\langle \log t_\star \rangle_L$, $\langle \log Z_\star \rangle_L$ e $\log M/L_r$ com uma precisão que varia entre 0,1 dex e 0,2 dex.
- Linhas de emissão: O resultado mais importante e original deste trabalho é que com a técnica apresentada podemos medir linhas de emissão com precisão maior que qualquer método direto de medida existente. Modelos que medem propriedades físicas de galáxias existem (por exemplo, Gallazzi et al. (2005) e possíveis adaptações do STARLIGHT) e estão sendo aprimoradas, mas não existe nada parecido em relação a medidas de linhas de emissão. Pode-se até criar bases de modelos incluindo emissão nebular, porém estas não teriam tanta confiabilidade como bases de populações estelares simples. Medir linhas e razões de linhas de emissão com a precisão mostrada na seção 5.3 ($\sigma(\Delta \log W_{H\alpha})$ e $\sigma(\Delta \log [N II]/H\alpha) \approx 0,15$ dex) nos permite diferenciar galáxias de formação estelar intensa de núcleos ativos, histórias de formação estelar específica e até metalicidade nebular com os índices de linhas

$$[\text{O III}]/[\text{N II}] \text{ e } \frac{[\text{O II}]+[\text{O III}]}{\text{H}\beta}.$$

6.2 Perspectivas

Com os resultados que apresentamos na seção 5, estamos convencidos que este projeto tem um grande potencial. Como projeto de doutoramento, planejamos seguir este trabalho com os próximos passos resumidos nos seguintes itens:

- Realizar mais testes com outras bases construídas de outras formas e aferir que tipo de base é mais adequada a cada aplicação;
- Substituir as funções deltas de Dirac que representam cada propriedade p de cada galáxia da base por distribuições de probabilidade associadas à esta propriedade. Ou seja, temos que substituir a função $P(p|h_j) = \delta(p - p_j)$ por uma função distribuição de probabilidades realista, como, por exemplo, a função gaussiana. Não esperamos que este aprimoramento modifique de modo significativo nossos resultados, porém representa um requinte formal fácil de ser implementado (ainda que computacionalmente dispendioso);
- Não foram incluídos Quasares nem Seyferts do tipo 1 em nossa amostra geral e, portanto, tampouco as incluímos em nossas bases de referência e amostras de objetos. Em um estudo mais completo, estes tipos de galáxias devem ser incluídos;
- O passo mais importante (e também o mais trabalhoso) será incluir o *redshift* na análise. Por simplicidade, e para focar nosso trabalho no método em si, optamos por trabalhar com espectros sempre no referencial de repouso ($z = 0$), mas na prática as galáxias serão observadas entre $z = 0$ e ~ 1 (no caso do JPAS). Pode-se imaginar que z (ou sua PDF) será dado pelos outros membros da colaboração que o medirão com grande precisão para seus fins cosmológicos ou então incorporar z como uma propriedade a mais a ser determinada em nossas análises. Uma dificuldade maior é que, em *redshifts* maiores, as galáxias podem ser muito diferentes. A física do gás, principalmente, pode ser muito diferente (Stasińska 2008), não sabemos se é possível usar uma base construída em baixos *redshifts* para uma amostra em altos. A maneira para compensar esta perda da base de referência ainda não está clara. Estudos detalhados neste sentido devem ser realizados.

- O trabalho teórico acima descrito é interessante e importante, mas a prova real do método seria a aplicação a dados reais. Apesar de termos centrado nossas simulações neste trabalho ao JPAS, podemos adaptar a metodologia para qualquer sistema fotométrico. Assim, do ponto de vista prático podemos enumerar os próximos passos como a aplicação aos dados do ALHAMBRA que já tem fotometria em 20 filtros de largura média de 300 \AA disponível e, indo mais além, aplicar o método com os dados fotométricos do próprio SDSS (apenas 5 filtros de 1000 \AA). Naturalmente, para o SDSS, esperamos um resultado muito pior que para o ALHAMBRA, mas esta aplicação pode ser vista como um exercício que pode ser útil para nos dizer sobre o que pode ser recuperado com outros *surveys* similares que estão por vir (com destaque para o LSST).

Referências Bibliográficas

- Asari, N. V., Cid Fernandes, R., Stasińska, G., Torres-Papaqui, J. P., Mateus, A., Sodré, L., Schoenell, W., & Gomes, J. M. 2007, *MNRAS*, 381, 263
- Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Ž., Lupton, R. H., Nichol, R. C., & Szalay, A. S. 2004, *ApJ*, 600, 681
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, *PASP*, 93, 5
- Bell, E. F., Wolf, C., Meisenheimer, K., Rix, H.-W., Borch, A., Dye, S., Kleinheinrich, M., Wisotzki, L. et al. 2004, *ApJ*, 608, 752
- Benítez, N. 2000, *ApJ*, 536, 571
- Benítez, N., Moles, M., Aguerri, J. A. L., Alfaro, E., Broadhurst, T., Cabrera-Caño, J., Castander, F. J., Cepa, J. et al. 2009, *ApJ*, 692, L5
- Brinchmann, J., Charlot, S., White, S. D. M., Tremonti, C., Kauffmann, G., Heckman, T., & Brinkmann, J. 2004, *MNRAS*, 351, 1151
- Bruzual, G. & Charlot, S. 2003, *MNRAS*, 344, 1000
- Chilingarian, I., Prugniel, P., Sil'Chenko, O., & Koleva, M. 2007, in *IAU Symposium*, Vol. 241, *IAU Symposium*, ed. A. Vazdekis & R. F. Peletier, 175–176
- Cid Fernandes, R., Asari, N. V., Sodré, L., Stasińska, G., Mateus, A., Torres-Papaqui, J. P., & Schoenell, W. 2007, *MNRAS*, 375, L16
- Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska, G., & Gomes, J. M. 2005, *MNRAS*, 358, 363
- Cid Fernandes, R., Stasińska, G., Mateus, A., & Vale Asari, N. 2011, *MNRAS*, 413, 1687
- Cid Fernandes, R., Stasińska, G., Schlickmann, M. S., Mateus, A., Vale Asari, N., Schoenell, W., & Sodré, L. 2010, *MNRAS*, 403, 1036
- Driver, S. P. 2011, *ArXiv e-prints*
- Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, *MNRAS*, 362, 41

- Kauffmann, G., Heckman, T. M., White, S. D. M., Charlot, S., Tremonti, C., Brinchmann, J., Bruzual, G., Peng, E. W. et al. 2003, *MNRAS*, 341, 33
- Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, *ApJ*, 556, 121
- Koleva, M., Prugniel, P., Bouchard, A., & Wu, Y. 2009, *A&A*, 501, 1269
- Liang, Y. C., Hammer, F., Yin, S. Y., Flores, H., Rodrigues, M., & Yang, Y. B. 2007, *A&A*, 473, 411
- Mateus, A., Sodré, L., Cid Fernandes, R., & Stasińska, G. 2007, *MNRAS*, 374, 1457
- Mateus, A., Sodré, L., Cid Fernandes, R., Stasińska, G., Schoenell, W., & Gomes, J. M. 2006, *MNRAS*, 370, 721
- Ocvirk, P., Pichon, C., Lançon, A., & Thiébaud, E. 2006, *MNRAS*, 365, 74
- Panther, B., Heavens, A. F., & Jimenez, R. 2003, *MNRAS*, 343, 1145
- Panther, B., Jimenez, R., Heavens, A. F., & Charlot, S. 2007, *MNRAS*, 378, 1550
- Peeples, M. S., Pogge, R. W., & Stanek, K. Z. 2009, *ApJ*, 695, 259
- Seife, C. 2005, *Science*, 309, 78
- Shane, C. D., Wirtanen, C. A., & Steinlin, U. 1959, *AJ*, 64, 197
- Sivia, D. & Skilling, J. 2006, *Data Analysis: A Bayesian Tutorial* (Oxford, USA: Oxford University Press)
- Stasińska, G. 2008, in *IAU Symposium*, Vol. 255, *IAU Symposium*, ed. L. K. Hunt, S. Madden, & R. Schneider, 375–380
- Stasińska, G., Cid Fernandes, R., Mateus, A., Sodré, L., & Asari, N. V. 2006, *MNRAS*, 371, 972
- Stasińska, G., Vale Asari, N., Cid Fernandes, R., Gomes, J. M., Schlickmann, M., Mateus, A., Schoenell, W., & Sodré, Jr. Seagal Collaboration, L. 2008, *MNRAS*, 391, L29
- Tojeiro, R., Heavens, A. F., Jimenez, R., & Panther, B. 2007, *MNRAS*, 381, 1252
- Tremonti, C. A., Heckman, T. M., Kauffmann, G., Brinchmann, J., Charlot, S., White, S. D. M., Seibert, M., Peng, E. W. et al. 2004, *ApJ*, 613, 898
- Vale Asari, N., Stasińska, G., Cid Fernandes, R., Gomes, J. M., Schlickmann, M., Mateus, A., & Schoenell, W. 2009, *MNRAS*, 396, L71
- Walcher, C. J., Böker, T., Charlot, S., Ho, L. C., Rix, H.-W., Rossa, J., Shields, J. C., & van der Marel, R. P. 2006, *ApJ*, 649, 692

Way, M. J. 2011, ArXiv e-prints

York, D. G., Adelman, J., Anderson, J. E., Anderson, S. F., Annis, J., Bahcall, N. A., Bakken, J. A., Barkhouser, R. et al. 2000, AJ, 120, 1579