

UNIVERSIDADE FEDERAL DE SANTA CATARINA  
Centro de Ciências Físicas e Matemáticas – CFM

SÍNTESE DE POPULAÇÕES ESTELARES EM  
GALÁXIAS NA ERA DAS GRANDES BASES DE  
DADOS: MÉTODOS E APLICAÇÕES \*

Jean Michel Silva de Miranda Gomes

Dissertação realizada sob orientação do Prof.  
Dr. Roberto Cid Fernandes Jr. e apresentada à Pós-  
Graduação de Física da UFSC em preenchimento par-  
cial dos requisitos para obtenção do título de Mestre  
em Física.

Florianópolis  
2005

---

\* Trabalho financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

## AGRADECIMENTOS

- Ao meu pai “Manuel” por sempre me apoiar em todos os momentos e por comprar as revistas e livros de meu interesse científico.
- À minha mãe “Helena” por sempre estar presente em todas as minhas tarefas de casa desde a infância durante todo o meu período escolar.
- À minha irmã “Mireille” por fazer companhia nas horas difíceis.
- Ao meus tios, César e Tina, por me acolherem como um filho durante dois anos de meus estudos na graduação em Física.
- Aos meus tios de Portugal que mesmo distantes sempre gostam de enviar alguma lembrança.
- Em memória à minha vó “Zoê” por sempre conversar e me aconselhar sobre os meus planos futuros.
- Aos meus familiares (tios e primos) por fazer de minha alegria a deles.
- À minha noiva “Mayanna Karla” por me ajudar a prosseguir durante estes dois anos de trabalho árduo e pelo amor dedicado a mim.
- À José Alves Facundo por sempre estar à disposição : “... para o que der e vier”.
- À Maria Filomena de Figueredo Facundo, como minha mãe de consideração, que sempre me trata como um filho.
- À Mayara e Max por me hospedarem em São Paulo sempre de braços abertos.
- Ao Harlano por sempre escutar as músicas que eu gosto.
- Ao meu orientador Roberto Cid Fernandes por se fazer presente em cada passo de minha dissertação.
- Aos professores Raymundo Baptista e Antônio Nemer Kanaan do grupo de astrofísica da UFSC por todos os seus ensinamentos.
- Ao nossos colaboradores de pesquisa: Laerte, Grazyna e Abílio pela grande contribuição a este trabalho.
- Aos colegas do Grupo de Astrofísica e do curso de Física da UFSC.

- Aos funcionários da Universidade Federal de Santa Catarina pelo empenho e dedicação a esta unidade de ensino superior.
- Ao CNPq por todo apoio financeiro e por acreditar no desenvolvimento científico do Brasil.

## Resumo

Síntese espectral é a decomposição de um espectro observado a partir da combinação linear de uma base de populações estelares simples (SSP) de diferentes idades e metalicidades de maneira que possamos avaliar a história da formação estelar e química da galáxia. Neste trabalho, estudamos a confiabilidade de nosso método de síntese espectral por meio de simulações e testes empíricos. Ambos demonstraram que o método é muito poderoso. A combinação dessa ferramenta de análise com a disponibilidade de grandes base de dados de espectros de galáxias abre um grande leque de linhas de pesquisa relativos ao processo que envolve a formação e evolução de galáxias.

Estimulados pelos excelentes ajustes obtidos para uma amostra de  $\sim 50000$  galáxias da “Sloan Digital Sky Survey”, começamos um projeto para investigarmos mais de perto resultados para galáxias elípticas, com o objetivo de determinar a história de formação estelar e enriquecimento químico dessas. Comparações cuidadosas entre espectro observado e sintetizado (modelado) revelaram resíduos sistemáticos em algumas bandas de absorção devido a elementos  $\alpha$  onde os resíduos aumentam com o aumento da dispersão de velocidades. Isto acontece porque a base espectral usada nos ajustes possui uma mistura solar de elementos.

Investigamos os efeitos dessa discrepância por meio de simulações. Para nosso desânimo, encontramos que esse mau ajuste produz efeitos sistemáticos e não negligenciáveis na estimação da idade metalicidade estelar média. Uma solução completa para esse “problema” requer uma base enriquecida com elementos  $\alpha$ .

## Abstract

Spectral synthesis is the decomposition of an observed spectrum in terms of a linear combination of a base of simple stellar populations of different ages and metallicities in order to evaluate the star formation and chemical histories of a galaxy. In this work, we study the reliability of the spectral synthesis method by means of simulations and empirical tests, both of which demonstrate that the method is very powerful. The combination of this tool with the availability of huge data bases of galaxy spectra opens a variety of research areas related to the processes involved in the formation and evolution of galaxies.

Stimulated by the excellent fits obtained for a sample of  $\sim 50000$  galaxies from Sloan Digital Sky Survey, we started a project to investigate more closely the results for elliptical galaxies, aiming to determine their star formation and chemical histories. Careful comparison between observed and synthetic spectra reveals systematic residuals in absorption features due to  $\alpha$ -elements, and that these residuals increase for increasing velocity dispersion. This occurs because the spectral base used in the fits have a solar mixtures of elements.

We investigated the effects of this mismatch by means of simulations. We find that this  $\alpha$  mismatch produces systematic and non-negligible biases in both of mean stellar age and metallicity estimates. A full solution to this "problem" requires an  $\alpha$ -enhanced base.

# SUMÁRIO

1. <i>Introdução</i> . . . . .	11
1.1 Síntese de Populações Estelares . . . . .	12
1.2 Síntese de Populações Estelares Semi-Empírica . . . . .	13
1.2.1 Breve Histórico . . . . .	13
1.3 Trabalhos Recentes e o Presente Trabalho . . . . .	14
1.4 Sobre esse Trabalho . . . . .	16
2. <i>Teste do Método: Simulações</i> . . . . .	17
2.1 Descrição do Código de Síntese: STARLIGHT03 . . . . .	17
2.1.1 Procedimento de Ajuste . . . . .	18
2.1.2 Expansão da $\chi^2$ . . . . .	19
2.2 Aplicações da Síntese e a Necessidade de Simulações . . . . .	21
2.3 Simulações: Confiabilidade do Método . . . . .	22
2.3.1 Base $\mathbf{N}$ - $\mathcal{B}_N$ . . . . .	26
2.3.2 Galáxias Teste . . . . .	26
2.4 Caracterização de uma Galáxia: Idade e Metalicidade Estelar Média . . . . .	33
3. <i>Aplicação a Galáxias da SDSS</i> . . . . .	38
3.1 Amostra da SDSS Limitada em Volume . . . . .	38
3.2 Resultados Gerais da Síntese Espectral para a Amostra . . . . .	39
3.2.1 Extinção Estelar ( $A_V$ ) versus Extinção Nebular ( $A_V^{Balmer}$ ) . . . . .	39
3.2.2 Idade Estelar Média ( $\langle \log t_\star \rangle_L$ ) versus Largura Equivalente em $H\alpha$ e $D_n4000$ . . . . .	41
3.2.3 Massa Estelar versus Dispersão de Velocidade . . . . .	43
3.3 Considerações Sobre Estes Testes Empíricos . . . . .	44
4. <i>Síntese Espectral de Galáxias Elípticas: Problema dos Elementos <math>\alpha</math></i> . . . . .	46
4.1 Bases $\mathbf{O}$ , $\mathbf{P}$ & $\mathbf{Q}$ - $\mathcal{B}_O$ , $\mathcal{B}_P$ , $\mathcal{B}_Q$ . . . . .	48
4.2 A Distribuição de Idades Médias . . . . .	48
4.3 O “Triângulo das Bermudas” . . . . .	53
4.4 Simulações: O Triângulo das Bermudas . . . . .	53
4.5 Quem São Essas Galáxias? . . . . .	59
4.5.1 Resíduos Espectrais . . . . .	59
4.5.2 Extinção Negativa . . . . .	61
4.6 Galáxias Elípticas . . . . .	61
4.7 Simulações: $\alpha/Fe$ . . . . .	68
4.7.1 Configuração 1 ( $A_V$ livre) . . . . .	68
4.7.2 Configuração 2 ( $A_V \geq 0$ ) . . . . .	72

---

4.7.3	Configuração 3 ( $A_V \geq 0$ e elementos $\alpha$ mascarados) . . . . .	72
4.7.4	Sobre as Simulações: Configuração 1, 2 & 3 . . . . .	72
4.8	Possíveis cenários para enriquecimento $\alpha/Fe$ . . . . .	75
4.9	O Problema de Síntese para Galáxias Elípticas . . . . .	76
5.	<i>Considerações Finais e Perspectivas Futuras</i> . . . . .	77
5.1	Perspectivas . . . . .	78
	<i>Apêndice</i> . . . . .	81
A.	<i>Nomenclatura</i> . . . . .	82
B.	<i>Uma Formulação Probabilística para a Síntese de Populações Estelares</i> . .	83
B.1	Estatística Bayesiana . . . . .	83
B.2	Algoritmo de Metropolis . . . . .	85
B.2.1	“Burn-in” . . . . .	85
B.3	Formalismo Aplicado à Síntese . . . . .	86



## LISTA DE FIGURAS

2.1	Síntese espectral de uma galáxia “ <i>early type</i> ” da SDSS . . . . .	24
2.2	Síntese espectral de uma galáxia “ <i>late-type</i> ” da SDSS . . . . .	25
2.3	Base N: $N_{\star} = 45$ SSP com 15 idades e 3 metalicidades . . . . .	27
2.4	Diagrama idade estelar média versus metalicidade estelar média usado para gerar as galáxias teste . . . . .	29
2.5	Diagrama idade estelar média versus metalicidade estelar média ilustrando espectros médios em 16 caixas. . . . .	30
2.6	Espectro de uma galáxia teste gerada com parâmetros de entrada conhecidos e listados na tabela 2.2 e sua respectiva perturbação com um ruído gaussiano por meio de um espectro de erro característico. . . . .	31
2.7	Espectro de erro característico de galáxias da SDSS . . . . .	32
2.8	Parâmetros de entrada versus saída para 65 galáxias teste: $x_Y$ , $x_I$ , $x_O$ , $A_V$ , $\sigma_{\star}$ e $M_{\star}$ . . . . .	34
2.9	Valores de entrada contra saída para o diagrama idade estelar média e metalicidade estelar média para as simulações com 65 galáxias teste . . . . .	36
2.10	Resíduos da metalicidade estelar média contra resíduos da idade estelar média para simulações com 65 galáxias teste . . . . .	37
3.1	Exemplo de um espectro residual com linhas de emissão . . . . .	40
3.2	Relação entre extinção de Balmer e extinção estelar para nossa amostra da SDSS de galáxias do tipo “star-forming” . . . . .	42
3.3	Idade estelar média versus largura equivalente de $H\alpha$ e $D_n4000$ para sub-amostras de galáxias da SDSS . . . . .	43
3.4	Relação entre a $M_{\star}$ - $\sigma_{\star}$ para nossa amostra de galáxias da SDSS . . . . .	44
4.1	Histograma da idade estelar média ponderada pela massa e pelo fluxo para a amostra de 50362 galáxias da SDSS sintetizadas com a $\mathcal{B}_N$ . . . . .	47
4.2	Amostra completa de 50362 galáxias da SDSS no diagrama idade estelar média versus metalicidade estelar média . . . . .	49
4.3	<b>Base O:</b> $N_{\star} = 90$ SSP com 15 idades e 6 metalicidades . . . . .	50
4.4	<b>Base P:</b> $N_{\star} = 120$ SSP com 30 idades e 4 metalicidades . . . . .	51
4.5	<b>Base Q:</b> $N_{\star} = 152$ SSP com 38 idades e 4 metalicidades . . . . .	52
4.6	Histogramas da idade média ponderada pela massa e pelo fluxo para a amostra randômica de 1000 galáxias da SDSS sintetizadas com a $\mathcal{B}_N$ , $\mathcal{B}_O$ , $\mathcal{B}_P$ & $\mathcal{B}_Q$ . . . . .	54
4.7	Histograma do $D_n4000$ para a amostra de 50362 galáxias da SDSS sintetizadas com a $\mathcal{B}_N$ . . . . .	55

4.8	$D_n4000$ versus idades de populações estelares simples com uma das metalicidades dadas: $0.005 Z_\odot$ , $0.02 Z_\odot$ , $0.2 Z_\odot$ , $0.4 Z_\odot$ , $1.0 Z_\odot$ e $2.5 Z_\odot$ . . . . .	56
4.9	Amostra aleatória de 1000 galáxias da SDSS: idade média contra metalicidade média - Bases $\mathcal{B}_N$ , $\mathcal{B}_O$ , $\mathcal{B}_P$ & $\mathcal{Q}_N$ . . . . .	57
4.10	Amostra aleatória de 1000 galáxias da SDSS: comparação entre bases O, P, Q & N para os desvios em relação ao modelo ( $\Delta_\lambda$ ) . . . . .	58
4.11	Idade média e metalicidade estelar média de entrada versus saída e degenerescência entre idade e metalicidade para 85 galáxias teste geradas no <i>quadrante superior direito</i> do diagrama $\langle \log t_\star \rangle_L$ versus $\log \langle Z_\star \rangle_L$ . . . . .	60
4.12	Ajuste da Galáxia A . . . . .	62
4.13	Ajuste da Galáxia B . . . . .	63
4.14	Ajuste da Galáxia C . . . . .	64
4.15	Média dos espectros, resíduos e resíduos quadráticos para $\sim 2000$ galáxias elípticas . . . . .	65
4.16	Figura 4 do trabalho de Thomas, Maraston & Bender 2002 mostrando o enriquecimento de elementos $\alpha$ em função da dispersão de velocidade. . . . .	67
4.17	Resíduo das linhas de absorção Mg, CN e Na em função da dispersão de velocidade . . . . .	69
4.18	Galáxia teste com acréscimo de um resíduo médio nas bandas CN, Mg e Na - simulando um enriquecimento $\alpha$ . . . . .	70
4.19	Configuração 1 ( $A_V$ livre): Idade média e metalicidade média de entrada versus saída e degenerescência entre idade e metalicidade para 85 galáxias teste geradas no <i>quadrante superior direito</i> do diagrama $\langle \log t_\star \rangle_L$ versus $\log \langle Z_\star \rangle_L$ com acréscimo de um espectro de resíduo médio nas bandas CN, Mg e Na. . . . .	71
4.20	Configuração 2 ( $A_V \geq 0$ ): Idade média e metalicidade média de entrada versus saída e degenerescência entre idade e metalicidade para 85 galáxias teste geradas no <i>quadrante superior direito</i> do diagrama $\langle \log t_\star \rangle_L$ versus $\log \langle Z_\star \rangle_L$ com acréscimo de um espectro de resíduo médio nas bandas CN, Mg e Na. . . . .	73
4.21	Configuração 3: Idade média e metalicidade estelar média de entrada versus saída e degenerescência entre idade e metalicidade para 85 galáxias teste geradas no <i>quadrante superior direito</i> do diagrama $\langle \log t_\star \rangle_L$ versus $\log \langle Z_\star \rangle_L$ com acréscimo de um espectro de resíduo médio nas bandas CN, Mg e Na. . . . .	74

## LISTA DE TABELAS

2.1	Componentes do vetor $\vec{x}$ da $\mathcal{B}_N$ e suas respectivas idades e metalicidades estelares. . . . .	28
2.2	Parâmetros para uma galáxia teste gerada com a $\mathcal{B}_N$ . . . . .	33

# Capítulo 1

## Introdução

*“A common sense interpretation of the facts suggests that a superintellect has monkeyed with physics, as well as with chemistry and biology, and that there are no blind forces worth speaking about in nature.”*

**Fred Hoyle**

Galáxias são uma importante fonte de estudos para a cosmologia, pois a partir da determinação de suas características (massa, tamanho e história da formação estelar), podemos inferir como estas estruturas se formaram, evoluíram e estão evoluindo no universo. É interessante notar que existe uma seqüência morfológica bem definida de galáxias que, simplificando, pode ser colocada de acordo com três categorias, conforme proposto por Hubble (1926): elípticas, espirais e irregulares. Estas, por sua vez, possuem características bem diferentes entre si em termos de suas composições estelares, taxas de formação estelar, quantidade de gás, composições químicas, entre outras (Kenicutt 1998). Um método para analisar as propriedades das galáxias é através de seu espectro. Por não termos resolução suficiente para observar cada estrela de maneira individual (com exceção de galáxias do grupo local), em geral temos que nos contentar com um *espectro integrado*. Comumente, uma galáxia possui um espectro composto de: linhas de absorção mais contínuo devido à presença de estrelas, linhas de emissão devido ao gás morno, além de mais uma componente não estelar no caso de núcleos ativos.

Nesse trabalho, estamos interessados em estudar a história da formação estelar e evolução de galáxias a partir de observações. Adentrando em uma linha de pesquisa chamada síntese de populações estelares, podemos encontrar a resposta para solucionar o problema das mesmas.

Diferentes métodos de síntese foram desenvolvidos ao longo desses últimos 30 anos. Começaremos descrevendo a classificação e um pouco da história desses métodos.

## 1.1 Síntese de Populações Estelares

De forma geral, as diferentes técnicas podem ser classificadas em duas vertentes:

1. Síntese de Populações Estelares Evolutiva (SPEE)
2. Síntese de Populações Estelares Semi-Empírica (SPESE)

O primeiro desses métodos, introduzido por Tinsley (1968) e Spinrad (1972), trata de comparar os dados da galáxia com modelos que seguem a evolução temporal de um sistema estelar inteiro através da combinação de bibliotecas com caminhos evolutivos e espectros estelares em conjunto com a função de massa inicial (IMF<sup>1</sup>), história de formação estelar e evolução química. O espectro sintetizado  $M_\lambda$  é então comparado com o espectro real  $O_\lambda$ , e, se existirem diferenças significativas, os parâmetros de entrada serão reajustados. O grande problema dessa estratégia é a existência de muitas hipóteses envolvidas e informações incompletas ou imprecisas de nossos modelos de evolução estelar e atmosferas. No entanto, com o desenvolvimento de modelos cada vez mais precisos, esse problema está desvanecendo.

O segundo desses, desenvolvido pelos trabalhos pioneiros de Morgan (1956), Wood (1966) e Faber (1972), utiliza a informação contida nos espectros (contínuo e linhas de absorção ou emissão) como ponto de partida para se inferir a populações estelares existentes sendo, muitas vezes, por este motivo, denominada de uma técnica de “*inversão*”. Essa tem por objetivo reproduzir um conjunto de observações de uma dada galáxia ( $O_\lambda$ ) por meio da combinação linear de sistemas simples com características conhecidas, como estrelas individuais ou grupos de estrelas quimicamente homogêneas e coevas (aglomerados estelares).

Em ambos métodos, existem diferenças em relação ao conjunto de observáveis que são sintetizados. Tradicionalmente, a comparação entre dados e modelos é feita em termos de índices espectrais, como larguras equivalentes<sup>2</sup> de linhas de absorção com emissão e cores (razões de fluxo em diferentes  $\lambda$ 's). Os índices de Lick (Trager et al. 1998) e os definidos por Bica (1988) estão entre os mais usados para fins de síntese. Embora índices espectrais contenham muita informação, eles são no fundo uma versão resumida do espectro. Idealmente, a comparação entre observação e modelo deveria ser feita para todos os  $\lambda$ 's, pixel a pixel, como faremos nesse trabalho.

Seguiremos essa abordagem com alguns temperos de síntese de população estelar evolutiva.

---

<sup>1</sup> “Initial mass function”. Essa função informa qual a quantidade de estrelas existentes entre uma massa  $M$  e  $M + dM$ . Portanto, a forma da função de massa influencia a quantidade de material de uma galáxia que estará presa em estrelas de diferentes massas (Salpeter 1955).

<sup>2</sup> A largura equivalente ( $EW \equiv$  “equivalent width”) de uma determinada linha  $l$  é o fluxo de  $l$  ( $F(l)$ ) dividido pelo fluxo por comprimento de onda no contínuo em  $l$  ( $C_\lambda(l)$ ). Matematicamente, temos que a largura equivalente de  $l$  é  $EW(l) = F(l)/C_\lambda(l) = [\int_{\lambda_i}^{\lambda_f} (F_\lambda - C_\lambda) d\lambda] / (C_\lambda(l))$ , onde  $F_\lambda$  é o fluxo em cada comprimento de onda,  $C_\lambda$  é o fluxo no contínuo em cada comprimento de onda, e  $\lambda_i$  e  $\lambda_f$  é a janela no espectro que contém a linha  $l$ .

## 1.2 Síntese de Populações Estelares Semi-Empírica

Para um dado espectro observado  $O_\lambda$  de uma galáxia, o ideal seria obter a combinação linear dos grupos espectrais que melhor representem o objeto de estudo, ou seja, que nos forneça o melhor modelo  $M_\lambda$  em comparação com os dados observacionais ( $O_\lambda$ ). Estes grupos espectrais são o que chamaremos de *base* e o produto da síntese são as proporções de cada componente dessa ( $x_j$ ,  $j = 1, N_*$ , onde  $N_* \equiv$  número de elementos da base). Os  $x_j$  geram um vetor  $\vec{x}$ , conhecido como *vetor de populações*, pois contém a abundância relativa de cada componente existente na galáxia observada. Por sabermos as características de cada componente da base (sua idade  $t_*$  e metalicidade  $Z_*$ ), podemos inferir a história da formação estelar e enriquecimento químico da galáxia a partir de  $\vec{x}$ .

### 1.2.1 Breve Histórico

Os primeiros trabalhos de *SPESE* (Morgan 1956 e Wood 1966) utilizavam procedimentos de tentativa e erro para a determinação dos parâmetros. As cores dos objetos observados eram ajustadas a modelos de índice de cor sintéticos, mas falhavam na determinação das mesmas, pois muitas vezes encontravam-se abundâncias negativas de certos grupos espectrais que compõem a base (ou seja, alguns  $x_j < 0$ ). A partir desses desenvolvimentos iniciais, ficou claro que um desenvolvimento matemático mais rigoroso era necessário, levando-se em conta vínculos físicos (Faber 1972)<sup>3</sup>, como  $x_j > 0$  (vínculos de positividade) e  $\sum_j^{N_*} x_j = 1$  (vínculo de normalização). Outra característica do seu trabalho é a utilização de métodos de minimização, rejeitando os métodos por tentativa e erro. Nesse trabalho, foi adotado um critério de melhor ajuste onde minimiza-se a soma dos resíduos quadráticos percentuais em todas as bandas fotométricas. Se  $L_v$  é a luminosidade da galáxia no filtro  $v$ ,  $x_j$  é o número de estrelas do tipo  $j$  presentes.  $l_{vj}$  é a luminosidade no filtro  $v$  devido a estrelas do tipo  $j$ , e  $\Upsilon$  é o número total de filtros, portanto, a função a ser minimizada é  $Q = \sum_{v=1}^{\Upsilon} \omega_v (1 - \sum x_j l_{vj} / L_v)^2$ . Onde,  $\omega_v = \sigma_v^{-2}$ , é o peso e  $\sigma_v$  é o erro observacional em  $L_v$ .

Bica (1988) desenvolveu um método de síntese, que apesar de sua fundamentação matemática não rigorosa, produz bons resultados. Sabendo que as transições atômicas e moleculares das atmosferas estelares refletem as condições físicas e composições químicas, Bica utilizou larguras equivalentes como parte de seus observáveis, por estas serem boas traçadoras de idade e metalicidade de populações estelares. Seu algoritmo utiliza uma base de aglomerados observados composta de larguras equivalentes (9 linhas) para 35 aglomerados de idades e metalicidades conhecidas. Na verdade, os 35 espectros formam uma malha baseada em 63 espectros observados (Bica & Alloin 1986). Uma grande vantagem em usar uma base observacional é que a IMF, modelos de atmosferas e evolução estelar estão *implícitos* nesta, o que implica uma grande economia de hipóteses. Para a determinação das frações dos elementos da base, o código faz um teste de “todas” as possíveis combinações desses,

<sup>3</sup> Faber (1972); “... Os modelos resultantes eram fisicamente impossíveis porque eles continham números negativos de certos tipos de estrelas, ou tinham diagramas H-R astroficamente não razoáveis.”

que por essa razão é denominado procedimento de combinações diretas. A solução considerada é a média aritmética das soluções que reproduzem os observáveis dentro de uma janela de erro estipulada, sendo descartados os modelos com resíduo maior.

Sabemos que os elementos  $x_j$  devem variar entre 0 e 100%. Então, como temos uma amostragem uniforme no espaço de parâmetros, devemos fazer uma malha que quanto mais fina for a sua resolução,  $\Delta x_j$ , melhor resultado teremos, mas em contrapartida mais tempo de computação levaremos. Torna-se visível que esse método não tem muita eficiência pois muitos pontos têm pouca contribuição no cálculo dos posteriores (ver no apêndice seção B.1).

Posteriormente, outros avanços foram feitos. Schmidt et al. (1991) utilizaram um procedimento de multi-minimização. Schmidt também realizou testes com a base de 35 elementos de Bica mostrando que essa estava sobre-dimensionada (devido à dependência linear de suas componentes). Em 1994, Bica et al. também incluíram cores (razões de fluxo) e extinção nos observáveis sintetizados.

Os trabalhos de Pelat (1997 & 1998) contribuíram para desvendar alguns problemas em relação à solução obtida para o caso de termos mais variáveis que equações, ou seja, o número de observáveis é menor que o número de elementos da base (o sistema algébrico é indeterminado) e a isto denomina-se **degenerescência algébrica**. Esse é um problema que **não** teremos, já que utilizamos o espectro como um todo em nossa síntese.

Cid Fernandes et al. (2001) revisaram o problema de síntese introduzindo uma perspectiva probabilística, mas ainda continua utilizando como observáveis as larguras equivalentes e cores propostas por Bica. O problema de estimar o vetor  $\vec{x}$  torna-se um problema de amostragem dos parâmetros, ou seja, obter os posteriores  $P(x_j|D)$ ,  $j = 1, N_*$  que são as distribuições de probabilidades dos  $x_j$  em relação a um conjunto de dados observacionais. Além do vetor de populações, também são obtidas as estimativas da extinção na banda V ( $A_V$ ) incluindo essa como um novo parâmetro. O mapeamento dos posteriores é realizado através de um método de Monte Carlo, baseado em um algoritmo de Metropolis. A eficiência desse método de amostragem por importância é bem maior que dos métodos tradicionais de amostragem uniforme, pois vários valores para os parâmetros não contribuem na distribuição de probabilidades e assim minimiza-se tempo na execução do código.

### 1.3 Trabalhos Recentes e o Presente Trabalho

Nos dias de hoje, espectros de galáxias estão mais abundantes do que nunca. A **Sloan Digital Sky Survey (SDSS)**, foco de nossas aplicações da síntese em particular, está fornecendo um conjunto de dados com cerca de  $5 \times 10^5$  espectros de galáxias em uma faixa espectral que cobre 3800–9200 Å, com resolução de  $\frac{\lambda}{\Delta\lambda} \sim 1800$ . Por esse motivo, precisamos para o presente trabalho de um método eficiente de síntese que consiga lidar com estas grandes amostras de dados observacionais de boa qualidade, já que nosso entendimento sobre a constituição, formação e evolução de galáxias depende diretamente da forma como fazemos inferências a partir dos dados.

A maioria dos estudos das populações estelares em galáxias da SDSS tem sido feita com base em índices espectrais (Kauffman et al. 2003). Apenas dois grupos

estão utilizando todo o espectro  $O_\lambda$ , pixel a pixel; o “grupo de Edinburgo” (Panter et al. 2004), e o nosso (Cid Fernandes et al. 2005). Nossa estratégia é conceitualmente semelhante àquela desenvolvida por Heavens, Jimenez & Iahav (2000) e Panter, Heavens & Jimenez (2004), mas operacionalmente nosso código de síntese **STARLIGHT** (descrito em 2.1) difere do código deles, o **MOPED** (“Multiple Optimized Parameter Estimation and Data Compression”). Esse último apresenta um método que “a partir de um conjunto de dados  $\vec{O}$  (em nosso caso o espectro de uma galáxia) que inclui um sinal  $\vec{S}$  e um ruído  $\vec{N}$ , i.e.,  $\vec{O} = \vec{S} + \vec{N}$ , a idéia é encontrar vetores peso  $\vec{b}_j$  ( $j = 1, N \equiv$  número de parâmetros) tais que  $y_j = \vec{b}_j^t \vec{O}$  forneça a máxima informação possível sobre os parâmetros (história da formação estelar, metalicidade, entre outros). Estes números são então utilizados como o conjunto de dados para a análise probabilística”<sup>4</sup>. Os espectros são antes degradados para uma resolução de 20 Å a fim de reduzir ainda mais o tempo de computação.

Nosso método utiliza o espectro como um todo no processo de síntese. Além disso, nossa síntese espectral utiliza uma base de alta resolução espectral ( $\lambda/\Delta\lambda \sim 1000$ ), obtida a partir dos espectros de populações estelares simples de diferentes idades e metalicidades calculadas por Bruzual & Charlot (2003; BC03).

O presente estudo tem como objetivo principal em relação ao código de síntese **STARLIGHT**, testar a confiabilidade do método. Para tanto, seguimos três linhas nesse trabalho:

- **Simulações:** Procuramos obter o grau de confiança nos parâmetros encontrados pela síntese. As simulações são realizadas com um conjunto de galáxias teste (*fictícias*) geradas com propriedades conhecidas (histórias de formação estelar, dispersão de velocidades, extinção, entre outras). Tenta-se então, a partir da síntese, recuperar estes parâmetros de entrada. A comparação entre os valores para os parâmetros de entrada, previamente gerados, com os de saída, obtidos pelo código, nos leva a um resultado mais claro do nível de segurança de nosso método.
- **Aplicação/Teste Empírico:** Por meio da aplicação a galáxias da SDSS, tentamos recuperar resultados astrofisicamente razoáveis através de nosso método de síntese. Procuramos comparar estes resultados com medidas obtidas de maneiras independentes de nosso código, assim obtendo outra maneira de avaliarmos o nível de segurança do mesmo.
- **Estudo de problemas:** Após obtermos resultados excelentes, tanto em simulações quanto em teste empírico, procuramos examinar com mais detalhe dois resultados intrigantes que poderiam estar relacionados tanto a deficiências quanto a limitações em relação a nosso método de síntese e/ou base utilizada. Essa é uma etapa de análise mais minuciosa de nosso código para podermos saber com mais clareza qual(is) a(s) limitação(ões) do mesmo.

Assim sendo, os capítulos ficam divididos em:

---

<sup>4</sup> Panter, Heavens & Jimenez (2004)



- Capítulo 2 - Descrição matemática do problema de síntese de populações estelares e do código **STARLIGHT03**, utilizado ao longo desse trabalho. Uma série de simulações é realizada comparando-se parâmetros de entrada previamente conhecidos com parâmetros de saída obtidos pela síntese.
- Capítulo 3 - Apresentamos um teste empírico da síntese para uma amostra de 50362 galáxias da SDSS, mostrando algumas correlações com quantidades obtidas pela síntese e por outros métodos distintos, como por exemplo: extinção nebular contra extinção estelar, massa estelar contra dispersão de velocidades. Parte desses resultados foram publicados em Cid Fernandes et al. (2005; CF05) anexado no apêndice. Os resultados dos capítulos subsequentes são novos e ainda inéditos.
- Capítulo 4 - Novas simulações são realizadas para a verificação de resíduos em algumas bandas (Mg, CN e Na) em galáxias com altas idades e metalicidades estelares médias. Testes com novas bases são realizados para verificar a influência na escolha dessas nos resultados da síntese. As galáxias elípticas entram em cena por serem estas a apresentar maior resíduo nestas bandas. Simulações com galáxias teste que possuem altas idades e metalicidades estelares médias são realizadas com e sem enriquecimento de elementos  $\alpha$ .
- Capítulo 5 - Uma breve revisão dos principais resultados obtidos no decorrer desse trabalho. São resumidos os “problemas” encontrados na síntese e discutidas possíveis soluções. Por fim, mencionamos uma perspectiva futura em relação ao código de síntese com a utilização de uma nova base.

#### 1.4 Sobre esse Trabalho

Boa parte dos resultados que estão apresentados, principalmente no capítulo 3, são devidos à contribuição de nossos colaboradores: Laerte, Abílio e Grazyna que não podem deixar de ser mencionados nesse trabalho. Estes resultados, em conjunto com a primeira série de simulações realizadas (grande maioria feitas por mim), fazem parte de um artigo publicado, Cid Fernandes et al. (2005), que está anexado ao apêndice dessa dissertação.

Deve-se também mencionar que temos resultados completamente originais apresentados: a expansão do  $\chi^2$  (capítulo 2) que ainda não havia sido previamente documentada, e os problemas do capítulo 4.

# Capítulo 2

## Teste do Método: Simulações

*“Todos os indícios de um artifício, todas as manifestações de um design que existem no relógio existem também nas obras da natureza, com a diferença de que, na natureza, são maiores ou mais numerosos, e isso num grau que excede todo cômputo.”*

**Richard Dawkins**  
(Ditado Traduzido)

Descreveremos com mais detalhes como nosso código de síntese executa suas operações. Utilizamos uma técnica de SPESE com alguns temperos de SPEE (seção 2.1). A partir dos produtos diretos da síntese (principalmente o vetor de populações  $\vec{x}$ ), definimos duas quantidades que podem ser utilizadas para caracterizar galáxias: idade estelar média e metalicidade estelar média. Com estes desenvolvimentos, mostramos algumas simulações realizadas para testar a confiabilidade do método.

### 2.1 Descrição do Código de Síntese: *STARLIGHT03*

Nosso código de síntese utilizado nesse e no seguinte capítulos, o qual denominamos **STARLIGHT03**, mistura técnicas computacionais originalmente desenvolvidas para SPESE (Cid Fernandes et al. 2001) com ingredientes de modelos de SPEE (BC03). Basicamente é feito um ajuste ao espectro observado  $O_\lambda$  com a combinação linear de espectros de  $N_\star$  populações estelares simples (SSP  $\equiv$  “Simple Stellar Populations”), cada uma delas caracterizada por uma idade  $t_j$  e metalicidade  $Z_j$  ( $j = 1, N_\star$ ). Existem processos como absorção e espalhamento da luz que são provenientes do objeto observado. Estes ocorrem devido aos grãos de poeira no meio interestelar (ISM  $\equiv$  “Inter Stellar Medium”), o qual fazem parte do chamado fenômeno de “*extinção*”, muitas vezes denominado “*avermelhamento*” (a luz em comprimentos de onda no azul sofre mais com a extinção do que comprimentos de onda no vermelho). Sabemos também que as estrelas em galáxias não estão estáticas, portanto, devido ao efeito Doppler, temos uma suavização do espectro que deve ser levada em conta. Estrelas afastando-se da linha de visada sofrem um deslocamento para o vermelho e estrelas aproximando-se da mesma têm seus espectros deslocados para o azul. Como existem estrelas afastando-se e aproximando-se, o efeito total

é um alargamento das linhas de absorção. Esse alargamento é feito utilizando-se uma distribuição gaussiana  $G(v_*, \sigma_*)$  centrada em uma velocidade sistêmica  $v_*$  e com dispersão de velocidade estelar  $\sigma_*$ .

A partir dessas hipóteses, temos como criar nosso modelo  $M_\lambda$  para descrever o espectro de uma população composta:

$$M_\lambda = M_{\lambda_0} \left[ \sum_{j=1}^{N_*} x_j b_{j,\lambda} r_\lambda \right] \otimes G(v_*, \sigma_*) \quad (2.1)$$

onde,  $b_{j,\lambda} \equiv L_\lambda^{SSP}(t_j, Z_j) / L_{\lambda_0}^{SSP}(t_j, Z_j)$  é o espectro da  $j$ -ésima componente da SSP normalizado em um comprimento de onda  $\lambda_0$ , correspondendo a cada espectro  $L_\lambda^{SSP}(t_j, Z_j)$  uma população estelar de uma dada idade ( $t_j$ ) e metalicidade ( $Z_j$ ). O avermelhamento, devido à poeira, é levado em conta com  $r_\lambda \equiv 10^{-0.4(A_\lambda - A_{\lambda_0})}$ .  $A_\lambda$  é a extinção no comprimento de onda  $\lambda$ ;  $A_{\lambda_0}$  é a extinção no comprimento de onda de normalização, sendo a mesma extinção para todas as componentes  $x_j$ . Nesse trabalho utilizamos o avermelhamento na banda V (visível)  $A_V$ . Vemos também que  $M_{\lambda_0}$  funciona como um fator de normalização, sendo igual ao fluxo sintético no comprimento de onda de normalização  $\lambda_0$ .  $\vec{x}$  é o *vetor de populações* que, como foi mencionado, é o principal produto da síntese, servindo para retratar a história da evolução estelar e enriquecimento químico da galáxia. Cada  $x_j$  corresponde à contribuição percentual da componente  $j$  da base com idade  $t_j$  e metalicidade  $Z_j$  ao fluxo do modelo em  $\lambda_0$ .

### 2.1.1 Procedimento de Ajuste

O ajuste do modelo é feito através de um método de “*simulated annealing*” + “*algoritmo de Metropolis*” para evitar mínimos locais, ou seja, amostramos por uma seqüência de iterações, o espaço de parâmetros  $\zeta$  ( $\vec{x}$ ,  $A_V$ ) (utilizamos a extinção “ $A_V$ ” na banda V de visual neste trabalho), em busca de um mínimo global da função  $\chi^2$  dada por:

$$\chi^2 = \sum_{\lambda} [(O_\lambda - M_\lambda)\omega_\lambda]^2 \quad (2.2)$$

Onde,  $\omega_\lambda^{-1}$  é o erro em  $O_\lambda$ . Regiões ao redor de linhas de emissão, píxeis ruins ou resíduos de céu são mascarados através da atribuição  $\omega_\lambda \equiv 0$ . Em outras palavras, não faremos a comparação entre essa parte do modelo com o espectro observado por não fazer sentido (o efeito matemático é que na  $\chi^2$  estes termos mascarados não serão “*somados*”). Os píxeis que desviam mais do que 3 vezes o rms<sup>1</sup> entre  $O_\lambda$  e uma estimativa inicial de  $M_\lambda$  são descartados fazendo com que o peso  $\omega_\lambda$  seja nulo.

A principal diferença em relação à versão do código usada em Cid Fernandes et al. (2004b CF04) e Garcia-Rissman et al. (2005), que é usado em diferentes aplicações, é que o programa executa uma expansão em séries do termo de extinção ( $r_\lambda$ ) que permite uma computação muito mais rápida da  $\chi^2$  do que na equação 2.2. Esse artifício matemático é detalhado na seção seguinte. A leitura dessa não se faz

<sup>1</sup> “root mean square” - raiz quadrada da média dos quadrados dos valores

obrigatória, pois não acarreta perda de informações significativas ao entendimento da dissertação.

### 2.1.2 Expansão da $\chi^2$

Podemos dividir a equação 2.2 em três termos:

$$\chi^2 = \sum_{\lambda} \omega_{\lambda}^2 O_{\lambda}^2 - 2 \sum_{\lambda} \omega_{\lambda}^2 O_{\lambda} M_{\lambda} + \sum_{\lambda} \omega_{\lambda}^2 M_{\lambda}^2 \quad (2.3)$$

Designaremos cada um dos três termos por  $\chi_1^2$ ,  $\chi_2^2$ ,  $\chi_3^2$ , respectivamente. O primeiro termo,  $\chi_1^2 \equiv \sum_{\lambda} \omega_{\lambda}^2 O_{\lambda}^2$ , é sempre o mesmo para todas as iterações do código, ou seja, para quaisquer valores dos parâmetros  $\vec{x}$ ,  $A_V$ ,  $M_{\lambda_0}$ ,  $\sigma_{\star}$  e  $v_{\star}$ . O termo  $\chi_2^2$  pode ser escrito como:

$$\chi_2^2 \equiv -2 \sum_{\lambda} \omega_{\lambda}^2 O_{\lambda} \left( M_{\lambda_0} \sum_{j=1}^{N_{\star}} x_j b_{j,\lambda} r_{\lambda} \otimes G(v_{\star}, \sigma_{\star}) \right) \quad (2.4)$$

Na equação 2.1 temos que convoluir a gaussiana  $G(v_{\star}, \sigma_{\star})$  com a quantidade  $b_{j,\lambda} r_{\lambda}$ , pois essa possui termos dependentes em  $\lambda$ . Porém, na escala do filtro gaussiano  $G(v_{\star}, \sigma_{\star})$ , o termo em  $r_{\lambda}$  é aproximadamente constante pois a lei de extinção é uma função suave de  $\lambda$ . Conseqüentemente,  $[b_{j,\lambda} r_{\lambda}] \otimes G(v_{\star}, \sigma_{\star}) \sim [b_{j,\lambda} \otimes G(v_{\star}, \sigma_{\star})] r_{\lambda}$ .

Optamos, então, por utilizar  $b_{j,\lambda} \otimes G(v_{\star}, \sigma_{\star})$  e chamaremos essa quantidade de  $\gamma_{j,\lambda}$ . Definindo  $y_j \equiv x_j M_{\lambda_0}$  (termo que não depende de  $\lambda$ ), a equação 2.4 pode ser reescrita como:

$$\chi_2^2 \equiv -2 \sum_{\lambda} \omega_{\lambda}^2 O_{\lambda} \left( \sum_{j=1}^{N_{\star}} y_j \gamma_{j,\lambda} r_{\lambda} \right) \quad (2.5)$$

Podemos reescrever o termo de extinção,  $r_{\lambda}$ , da seguinte maneira:

$$r_{\lambda} = 10^{-0.4(A_{\lambda} - A_{\lambda_0})} = 10^{-0.4A_V(q_{\lambda} - q_{\lambda_0})} = 10^{R_{\lambda}A_V} = e^{\rho_{\lambda}A_V} \quad (2.6)$$

onde,  $R_{\lambda} \equiv -0.4(q_{\lambda} - q_{\lambda_0})$ ,  $q_{\lambda} = A_{\lambda}/A_V$  e  $A_V$  é a extinção na banda V. Fazendo uma pequena mudança de variáveis, obtemos  $r_{\lambda}$  em termos de uma exponencial na qual o argumento  $\rho_{\lambda} = \ln(10)R_{\lambda}$ . Fizemos estas alterações a fim de expandirmos com mais facilidade esse termo em uma série de Taylor, mas temos um problema. O argumento da exponencial tem que ser pequeno para garantir a validade dessa expansão, mas sabemos que  $\rho_{\lambda}A_V$  pode **não** ser um número muito pequeno ( $\ll 1$ ). Para minimizarmos esse efeito, fazemos com que ele se torne menor ao subtrairmos  $\bar{\rho}_{\lambda}A_V$ , onde  $\bar{\rho}_{\lambda} \equiv \eta = \sum_{\lambda} \rho_{\lambda} / \sum_{\lambda}$  (média dos  $\rho_{\lambda}$  sobre todos os  $\lambda$  do espectro) e assim podemos reescrever a equação 2.6:

$$r_{\lambda} = e^{\rho_{\lambda}A_V} = e^{(\rho_{\lambda}A_V + \eta A_V - \eta A_V)} = e^{(\eta A_V)} e^{(\beta_{\lambda} A_V)} = e^{(\eta A_V)} \sum_{n=0}^{\infty} \frac{(A_V \beta_{\lambda})^n}{n!} \quad (2.7)$$

onde,  $\beta_{\lambda} = (\rho_{\lambda} - \bar{\rho}_{\lambda})$ . Portanto, a partir desses desenvolvimentos, podemos escrever o segundo termo da  $\chi^2$  como:

$$\chi_2^2 \equiv \sum_{j=1}^{N_*} y_j e^{\eta A_V} \sum_{n=0}^{\infty} (A_V)^n \left( -\frac{2}{n!} \sum_{\lambda} \omega_{\lambda}^2 O_{\lambda} \gamma_{j,\lambda} (\beta_{\lambda})^n \right) \quad (2.8)$$

$$= \sum_{j=1}^{N_*} y_j e^{\eta A_V} \sum_{n=0}^{\infty} (A_V)^n H_1(n, j) \quad (2.9)$$

A função  $H_1(n, j) \equiv -\frac{2}{n!} \sum_{\lambda} \omega_{\lambda}^2 O_{\lambda} \gamma_{j,\lambda} (\beta_{\lambda})^n$  **não** depende do vetor de populações  $\vec{x}$  e nem da extinção  $A_V$ . Portanto, esse termo não precisa ser calculado toda vez que temos uma mudança nestes valores, que vão depender somente de  $v_*$  e  $\sigma_*$ .

De forma similar, o terceiro termo da  $\chi^2$  pode ser colocado sob a forma:

$$\chi_3^2 \equiv \sum_{\lambda} \omega_{\lambda}^2 M_{\lambda}^2 = \sum_{\lambda} \omega_{\lambda}^2 \left( \sum_{j=1}^{N_*} y_j \gamma_{j,\lambda} 10^{R_{\lambda} A_V} \right) \left( \sum_{k=1}^{N_*} y_k \gamma_{k,\lambda} 10^{R_{\lambda} A_V} \right) \quad (2.10)$$

$$\chi_3^2 \equiv \sum_{\lambda} \omega_{\lambda}^2 \sum_{j=1}^{N_*} \sum_{k=1}^{N_*} y_j y_k \gamma_{j,\lambda} \gamma_{k,\lambda} 10^{2R_{\lambda} A_V} \quad (2.11)$$

Utilizando os mesmos argumentos e procedimentos anteriores na expansão da  $\chi_2^2$ , obtemos:

$$\chi_3^2 \equiv \sum_{\lambda} \omega_{\lambda}^2 \sum_{j=1}^{N_*} \sum_{k=1}^{N_*} y_j y_k \gamma_{j,\lambda} \gamma_{k,\lambda} e^{2\eta A_V} \sum_{n=0}^{\infty} \frac{(2A_V)^n (\beta_{\lambda})^n}{n!} \quad (2.12)$$

$$= \sum_{j=1}^{N_*} \sum_{k=1}^{N_*} y_j y_k e^{2\eta A_V} \sum_{n=0}^{\infty} (2A_V)^n H_2(n, j, k) \quad (2.13)$$

Como no caso anterior, a função  $H_2(n, j, k) \equiv \sum_{\lambda} \frac{\omega_{\lambda}^2}{n!} \gamma_{j,\lambda} \gamma_{k,\lambda} (\beta_{\lambda})^n$  **não** depende do vetor de populações  $\vec{x}$  e da extinção  $A_V$  e, conseqüentemente, não necessita ser calculada toda vez que estes valores mudam.

Agora que conseguimos fazer as expansões necessárias, podemos expressar a função  $\chi^2$  para  $n'$  termos da expansão como:

$$\chi^2 \sim \chi_1^2 + \sum_{j=1}^{N_*} y_j e^{\eta A_V} \sum_{n=0}^{n'} (A_V)^n H_1(n, j) + \sum_{j=1}^{N_*} \sum_{k=1}^{N_*} y_j y_k \sum_{n=0}^{n'} e^{2\eta A_V} (2A_V)^n H_2(n, j, k) \quad (2.14)$$

Assim sendo, o código fica muito mais rápido em relação a equação 2.2, pois somente precisamos calcular alguns termos (dependendo da precisão desejada) e além disto não necessitamos calcular a todo momento as funções  $H_1(n, j)$  e  $H_2(n, j, k)$ , pois estas não dependem do vetor de populações e da extinção, mas somente de  $v_*$  e  $\sigma_*$ , que são calculados em algumas iterações por uma rotina diferente no programa através da minimização da função  $\chi^2$  para valores fixos de  $\vec{x}$  e  $A_V$ .

Quantitativamente, a grande vantagem dessa expansão é o chamado “speed up factor”<sup>2</sup> (s.u.f.), que pode ser definido como a razão entre o número de iterações do

---

<sup>2</sup> Fator de velocidade: determina quantas iterações a mais pode-se realizar comparando-se um algoritmo com outro semelhante.

código sem expansão do  $\chi^2$  ( $N_{isexp}$ ) dividido pelo número de iterações com expansão do  $\chi^2$  ( $N_{icexp}$ ), ou seja :

$$s.u.f. = \frac{N_{isexp}}{N_{icexp}} \quad (2.15)$$

O número de iterações sem a expansão pode ser inferido a partir da equação 2.2 como sendo o número de passos dados ( $N_{steps}$ ) no espaço de parâmetros vezes o número de comprimentos de onda ( $N_\lambda$ ) vezes a quantidade de elementos da base ( $N_\star$ ). Assim, temos que  $N_{isexp} \sim N_{steps} \times N_\lambda \times N_\star$ .

Para a expansão, admitindo-se que na equação 2.14 o termo dominante seja a somatória dupla, o número de iterações é da ordem de  $N_{steps}$  vezes  $N_\star^2$  vezes o número de vezes que calculamos o somatório em  $n'$ . O somatório em  $n'$  é praticamente constante, pois o calculamos somente quando  $A_V$ ,  $v_\star$  ou  $\sigma_\star$  mudam. Estes dois últimos parâmetros mencionados não variam a cada iteração, portanto somente precisamos calcular o somatório em  $n'$  quando a extinção  $A_V$  varia. Como o algoritmo modifica apenas um parâmetro por iteração, essa varia numa frequência de  $N_{steps}/(N_\star + 1)$  e, conseqüentemente, o número de vezes a ser calculado o somatório em  $n'$  é uma vez (para a primeira vez em que o código é acionado) mais a frequência de vezes em que a extinção varia. Portanto, como são  $n'$  termos na somatória de  $H_2(n, j, k)$ , temos que  $N_{icexp} \sim N_{steps} \times N_\star^2 + n' \times (1 + N_{steps}/[N_\star + 1])$ . Como temos uma matriz simétrica em relação à troca de  $y_j$  e  $y_k$ , somente precisamos calcular  $N_\star/2$  termos e podemos escrever  $N_{icexp}$  como  $N_{steps} \times (N_\star/2)^2 + n' \times 1 + N_{steps}/[(N_\star/2) + 1]$ . A equação 2.15 fica então:

$$s.u.f. \sim \frac{N_{steps} N_\lambda N_\star}{N_{steps} (N_\star/2)^2 + n' [1 + N_{steps}/(N_\star/2)]} \quad (2.16)$$

Assim, temos que:

$$s.u.f. \sim \frac{N_\lambda N_\star}{(N_\star/2)^2 + n' [1/N_{steps} + 1/(N_\star/2)]} \quad (2.17)$$

Como o número de comprimentos de onda é da ordem de 4000 pontos, a quantidade de elementos da base é da ordem de 45 e o número de termos computados na somatória em  $n' \sim 10$ , temos que a equação 2.17 implica:

$$s.u.f. \sim 4 \left( \frac{N_\lambda}{N_\star} \right) \sim 350 \quad (2.18)$$

Conseguimos um fator  $s.u.f. \sim 350$ . Na prática, observamos um ganho um pouco menor ( $s.u.f. \sim 250$ ), mas ainda assim substancial. Esse é o fato que justifica a expansão em séries da  $\chi^2$ .

## 2.2 Aplicações da Síntese e a Necessidade de Simulações

Temos em mãos um método de síntese que pode ter aplicações diversas como:

- **Medir as linhas de emissão do espectro  $O_\lambda$ :** esta é uma tarefa complicada já que linhas de emissão possuem contaminação devido a populações estelares,

por exemplo: absorção em baixo da emissão em  $H\alpha$ . Com o código de síntese, obtemos um espectro sintético  $M_\lambda$  que possui linhas de absorção e contínuo estelares. O espectro residual ( $O_\lambda - M_\lambda$ ) é então o espectro “puro” das linhas de emissão do gás que assim são facilmente medidas. Esse procedimento é empregado no estudo da Kauffmann et al. 2003 (K03a, b e c), por exemplo.

- **Obter dispersão de velocidades estelares ( $\sigma_*$ ):** regiões relativamente limpas do espectro que não estão contaminadas por linhas de emissão são boas traçadoras de dispersão de velocidade. É por esse motivo que existem estudos como os de Vega (2004) e Garcia-Rissman et. al. (2005) sobre o tripleto do Cálcio (CaT), em  $\lambda\lambda$  8498, 8542, 8662 Å, que é utilizado para este propósito. Estes trabalhos foram realizados com o código **STARLIGHT02**<sup>3</sup>, mostrando assim sua grande versatilidade.
- **Inferir história da formação estelar e enriquecimento químico:** a reconstrução da história evolutiva de uma galáxia é obtida a partir de  $\vec{x}$  (vetor de populações). Exemplo desse tipo de aplicação está em CF04, que apresenta uma análise dessas populações existentes na região central ( $R \sim 200$  pc) para galáxias do tipo Seyfert 2 (Nesse estudo se adicionam uma lei de potência à base, para representar a componente não estelar).

Como o código destina-se a vários estudos, o nível de segurança em relação ao mesmo deve ser avaliado. Faremos isto por meio de simulações (seção 2.3) e um teste empírico com galáxias da SDSS (capítulo 3).

### 2.3 Simulações: Confiabilidade do Método

As figuras 2.1 e 2.2 ilustram alguns ajustes obtidos com nosso método de síntese. Vemos que os ajustes são “perfeitos”, mas estes não garantem a confiança plena do método devido a uma série de degenerescências envolvidas no processo de síntese. Estas fazem com que alguma das componentes da base, a princípio linearmente independentes, tornem-se uma combinação linear de demais componentes. Assim, diferentes componentes da base podem gerar o mesmo bom ajuste, indicando uma **multiplicidade de soluções** para a síntese. Essa multiplicidade aparece graças à combinação de três fatores:

1. **Degenerescência algébrica:** no primeiro capítulo desse trabalho, fizemos referência aos trabalhos pioneiros de síntese, que utilizavam somente determinados observáveis (como largas equivalentes e cores) para a síntese. Se tivermos menos observáveis que elementos da base, geramos um problema matemático de mais variáveis que equações. Esse é um problema **não** encontrado em nosso método porque utilizamos todo o espectro  $O_\lambda$ , que excede o número de parâmetros. Mesmo para bases com muitas componentes, temos que  $N_\lambda \gg N_*$ .

<sup>3</sup> A diferença desse para o **STARLIGHT03** é a expansão em série da  $\chi^2$  como explicado na subseção 2.1.2.

2. **Degenerescências intrínsecas de populações estelares:** existem efeitos de compensação em populações estelares ainda não muito bem entendidos que podem ser minimizados ao utilizarmos todo o espectro  $O_\lambda$ . O efeito mais claro é o da **degenerescência entre idade e metalicidade**, que é a incapacidade de diferenciação entre populações estelares que encontram-se com o espectro mais avermelhado devido a populações com idades mais velhas e menos metálicas ou por estas serem mais metálicas e mais novas.
3. **Dependência linear estatística:** com o aumento dos erros, as diferenças entre elementos da base tornam-se insignificantes, ou seja, componentes de propriedades espectrais similares são confundidas por estas serem muito semelhantes na presença de ruído.

A discussão anterior mostra claramente a necessidade de avaliarmos o grau de confiança nos parâmetros envolvidos na síntese. Conforme dito, os ajustes espectrais dependem de  $N_\star + 3$  parâmetros:  $N_\star - 1$  componentes de  $\vec{x}$ ,  $M_{\lambda_0}$ ,  $A_V$ , e os dois parâmetros cinemáticos  $v_\star$  e  $\sigma_\star$ . A confiabilidade do método é melhor estudada por meio de simulações. A partir de espectros fictícios, gerados com parâmetros conhecidos, adicionando-se ruído gaussiano, podemos examinar e comparar resultados de entrada com os de saída.

Uma série de simulações já havia sido feita em CF04, onde os espectros utilizados estão em um intervalo entre 3500–5200 Å e a base é composta de  $N_\star = 20$  elementos. Essa concentrou-se na recuperação do  $\vec{x}_{entrada}$  pelo método. Resultados de maior importância foram:

- Na ausência de ruído, o método recupera com grande precisão os valores de entrada de  $\vec{x}$ .
- Na presença de ruído, as componentes individuais  $x_j$  desviam dos valores de entrada. Essencialmente o que ocorre é o fator de degenerescência (3) da dependência linear estatística. Esse é um problema comum na síntese espectral, já mencionado em outros trabalhos como Panter et al. (2004), Tremonti (2002) e Cid Fernandes et al. (2001).

Após a determinação da origem das discrepâncias observadas nos valores de entrada e saída, não devemos acreditar numa descrição final do método em relação à distribuição de suas componentes, mas sim fazer um agrupamento em relação a componentes espectralmente similares e assim trabalharmos com quantidades mais robustas. Na realidade, o que estamos fazendo é condensar o vetor de populações projetando-o num espaço de menos dimensões, deixando de lado os detalhes de todas as idades e metalicidades envolvidas no problema, fornecendo um resultado mais confiável.

Uma nova série de simulações, diferentes daquelas em CF04, e melhor elaboradas era necessária a fim de averiguarmos alguns aspectos em relação aos parâmetros fornecidos pela síntese.



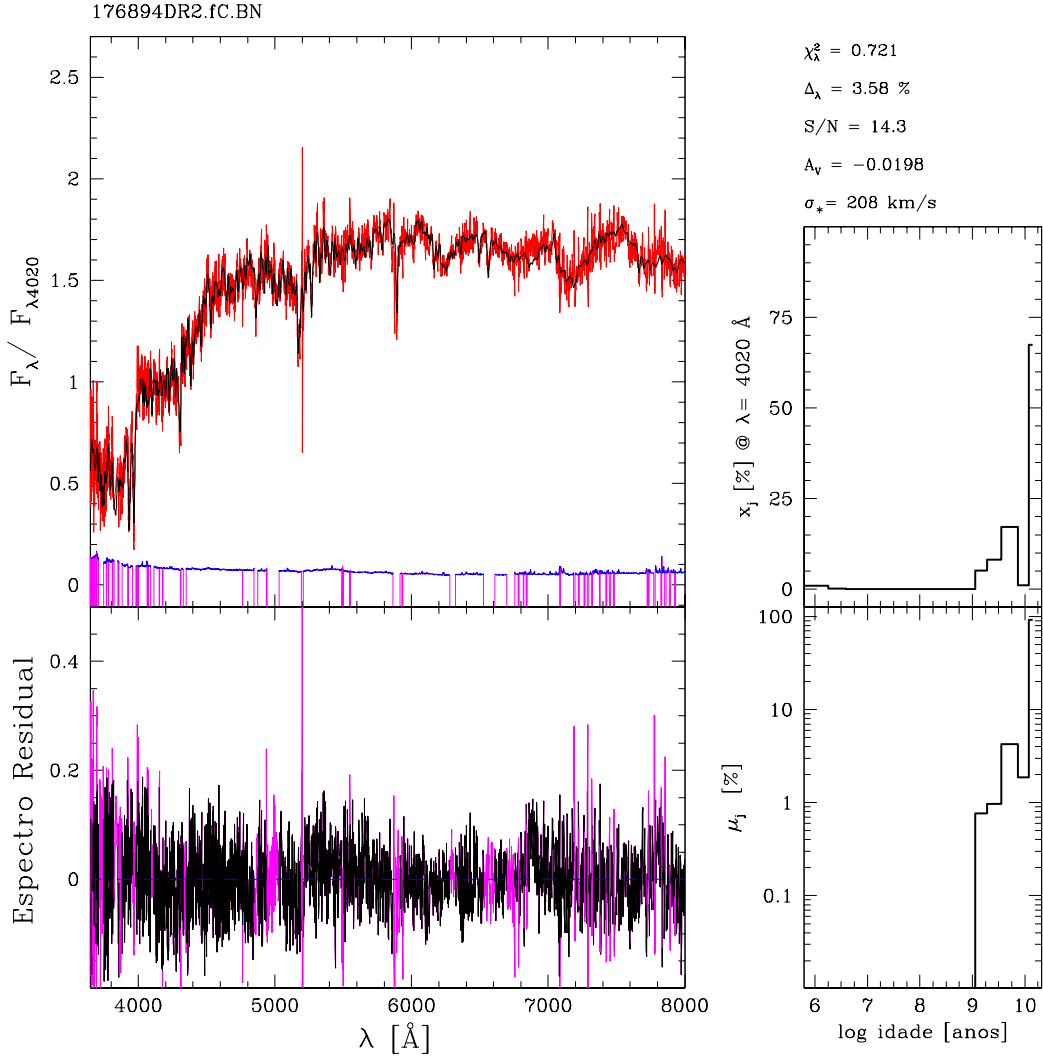


Fig. 2.1: Síntese espectral de uma galáxia “*early type*” da SDSS. Acima à esquerda temos o espectro: observado  $O_\lambda$  (em vermelho), modelado  $M_\lambda$  (em preto) e o erro (linha azul). Regiões marcadas (em magenta) no espectro de erro indicam píxeis ruins (dados por “*flags*” da SDSS) ou por janelas onde contenham linhas de emissão, ambos mascarados no ajuste. Abaixo à esquerda: espectro residual. Regiões mascaradas estão representadas pela linha magenta. À direita, vemos as frações de cada componente da base que correspondem ao fluxo (em cima) e que correspondem à massa (em baixo) em função do logaritmo da idade. Algumas propriedades estão listadas acima desse gráfico.  $\chi_\lambda^2$  é o  $\chi^2$  reduzido, e  $\Delta_\lambda$  é a diferença relativa média entre modelo e espectro observado ( $\overline{M_\lambda - O_\lambda}$ );  $S/N$  refere-se ao sinal ruído na região em torno do comprimento de normalização  $\lambda_0 = 4020 \text{ \AA}$ .

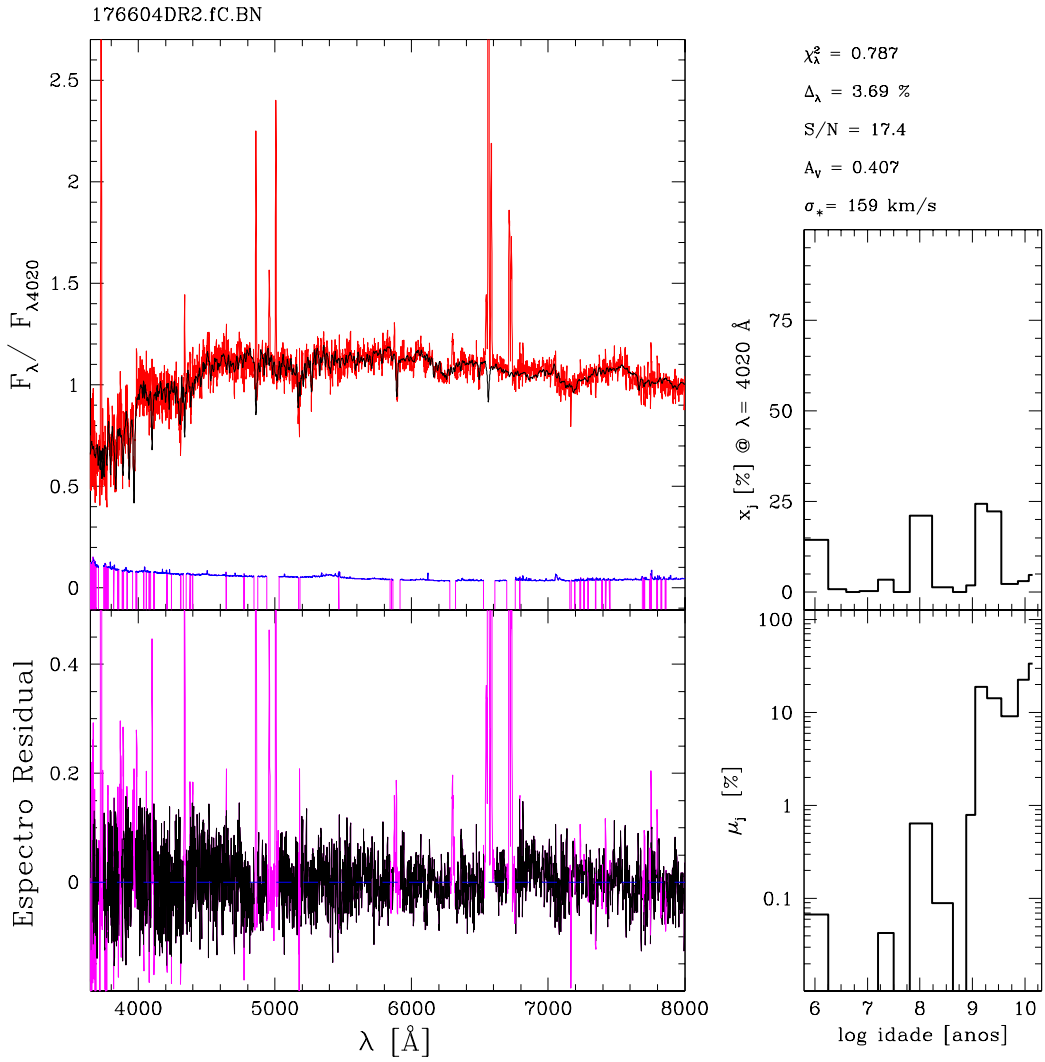


Fig. 2.2: Mesmo gráfico da figura 2.1, mas aqui estamos mostrando uma galáxia “late-type”.

### 2.3.1 Base $N$ - $\mathcal{B}_N$

Para a seqüência de simulações descritas a seguir, utilizamos uma base de  $N_\star = 45$  populações estelares simples com 15 idades variando entre  $10^6$  e  $1.3 \times 10^{10}$  anos e 3 metalicidades:  $Z = 0.2, 1.0$  e  $2.5Z_\odot$ . Os espectros mostrados na figura 2.3 são computados com a biblioteca STELIB (Le Borgne et al. 2003), “Padova 1994 tracks”, e a função de massa inicial de Chabrier (2003). Para detalhes desses cálculos veja BC03.

Cada componente  $x_j$  corresponde a uma dada idade ( $t_j$ ) e metalicidade ( $Z_j$ ), distribuídas conforme a tabela 2.1. Essa base é a que iremos chamar daqui para frente de “ $\mathcal{B}_N$ ”.

### 2.3.2 Galáxias Teste

Devido ao nosso interesse em modelar galáxias da SDSS, seguimos aqui com simulações que possuem grande parte das características desse conjunto de dados. Existem basicamente três diferenças principais em relação às simulações antigas (CF04):

- a faixa espectral é maior: 3650–8000 Å;
- a nova base é maior ( $\mathcal{B}_N$ );
- introdução de um espectro de erro mais realístico (figura 2.7).

O primeiro conjunto de simulações foi feito para galáxias teste geradas aleatoriamente a partir do diagrama idade estelar média e metalicidade estelar média (idade e metalicidade médias são definidas posteriormente; para o presente momento podemos considerar como quantidades que caracterizam uma dada galáxia como um todo). Essas foram geradas utilizando-se a média de  $\bar{x}$ ,  $A_V$  e  $\sigma_\star$  de galáxias reais que caíram dentro de 65 caixas no diagrama idade estelar média versus metalicidade estelar média (figura 2.4). Podemos ver uma ilustração de espectros médios em 16 caixas escolhidas no diagrama idade média versus metalicidade média na figura 2.5.

Foram realizadas 20 versões perturbadas de cada espectro sintético para cinco níveis de ruído:  $S/N = 5, 10, 15, 20, 30$  em torno de  $\lambda_0 = 4020$  Å. Podemos ver um exemplo de uma galáxia teste (figura 2.6, espectro em preto) gerada com parâmetros de entrada mostrados na tabela 2.2 e perturbada com  $S/N = 20$  (espectro em vermelho).

Em vez de usarmos um espectro de erro plano, em cada  $\lambda$  foi assumido que o erro seguia uma distribuição gaussiana com amplitude obtida através de um espectro de erro médio para galáxias da SDSS, escalonando-o em  $\lambda_0$  de forma a obter o  $S/N$  desejado. Esse erro  $\sigma_\lambda$  (figura 2.7) decresce por um fator  $\sim 3$  entre 3650–6200 Å e aumenta por um fator de  $\sim 1.5$  em direção a 8000 Å. Uma máscara é utilizada durante o processo de síntese ao redor de linhas de emissão: [O II] $\lambda\lambda$ 3726,3729, [Ne III] $\lambda$ 3869, H $\epsilon$ , H $\delta$ , H $\gamma$ , H $\beta$ , [O III] $\lambda\lambda$ 4959,5007, [He I] $\lambda$ 5876, [O I] $\lambda$ 6300, [N II] $\lambda\lambda$ 6548,6583, H $\alpha$ , [S II] $\lambda\lambda$ 6717,6731 e uma linha de absorção NaD $\lambda$ 5890 devido ao meio interestelar.

De modo geral, estas simulações confirmam o que já havia sido dito antes, que as componentes individuais  $x_j$  são muito incertas. Devemos então seguir para uma

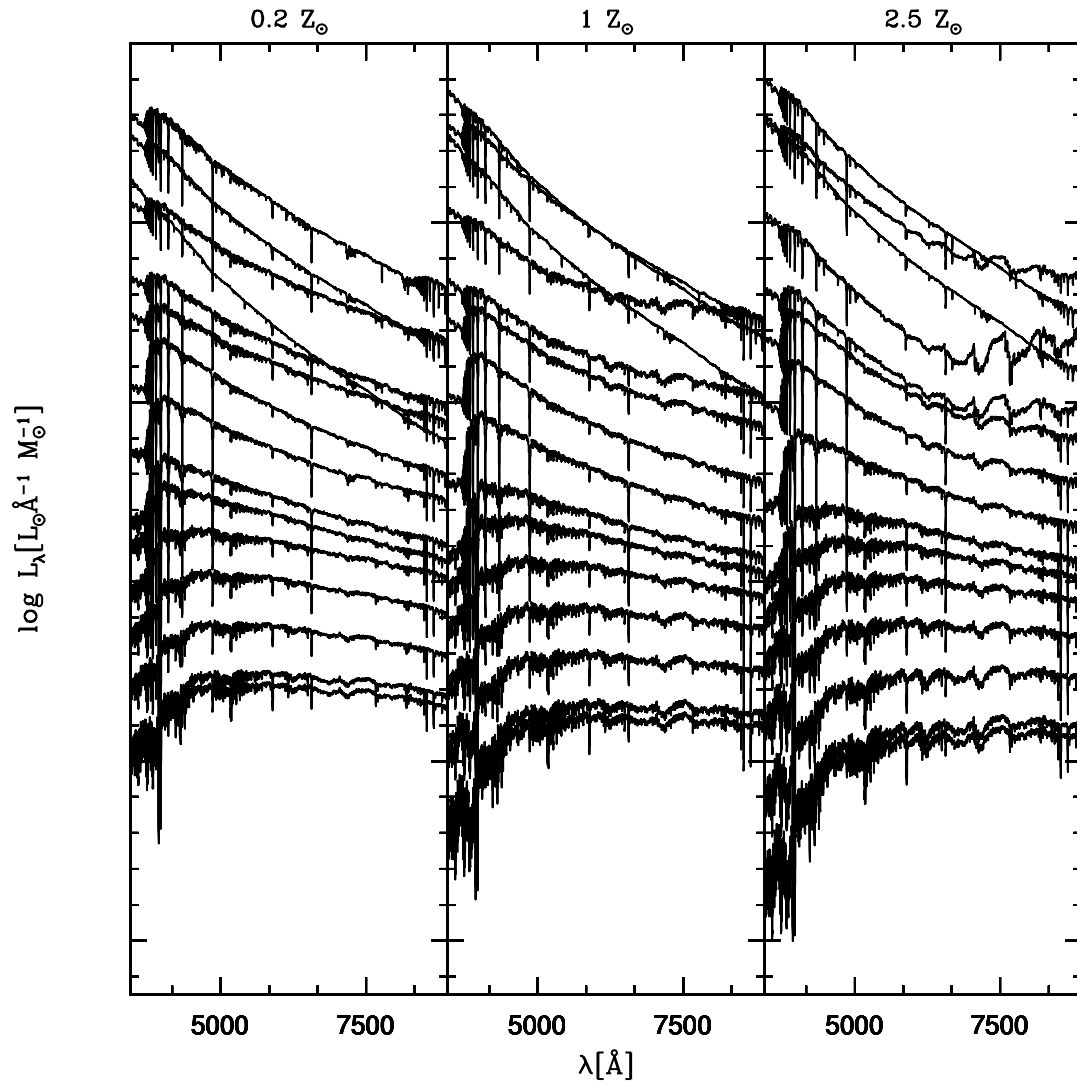


Fig. 2.3: Base N: Espectros de 45 populações estelares simples usados na síntese espectral (de BC03). A base compreende 3 diferentes metalicidades:  $Z = 0.2, 1.0$  e  $2.5 Z_{\odot}$ , e 15 idades: De cima para baixo,  $t = 0.001, 0.00316, 0.00501, 0.01, 0.02512, 0.04, 0.10152, 0.28612, 0.64054, 0.90479, 1.434, 2.5, 5, 11$  e  $13$  Ganos.

Base N - $\mathcal{B}_N$		
Componente da Base	Idade Estelar $t_*$	Metalicidade Estelar $Z_*$
$x_1$	0.001	0.2
$x_2$	0.00316	0.2
$x_3$	0.00501	0.2
$x_4$	0.01	0.2
$x_5$	0.02512	0.2
$x_6$	0.04	0.2
$x_7$	0.10152	0.2
$x_8$	0.28612	0.2
$x_9$	0.64054	0.2
$x_{10}$	0.90479	0.2
$x_{11}$	1.434	0.2
$x_{12}$	2.5	0.2
$x_{13}$	5	0.2
$x_{14}$	11	0.2
$x_{15}$	13	0.2
$x_{16}$	0.001	1.0
$x_{17}$	0.00316	1.0
$x_{18}$	0.00501	1.0
$x_{19}$	0.01	1.0
$x_{20}$	0.02512	1.0
$x_{21}$	0.04	1.0
$x_{22}$	0.10152	1.0
$x_{23}$	0.28612	1.0
$x_{24}$	0.64054	1.0
$x_{25}$	0.90479	1.0
$x_{26}$	1.434	1.0
$x_{27}$	2.5	1.0
$x_{28}$	5	1.0
$x_{29}$	11	1.0
$x_{30}$	13	1.0
$x_{31}$	0.001	2.5
$x_{32}$	0.00316	2.5
$x_{33}$	0.00501	2.5
$x_{34}$	0.01	2.5
$x_{35}$	0.02512	2.5
$x_{36}$	0.04	2.5
$x_{37}$	0.10152	2.5
$x_{38}$	0.28612	2.5
$x_{39}$	0.64054	2.5
$x_{40}$	0.90479	2.5
$x_{41}$	1.434	2.5
$x_{42}$	2.5	2.5
$x_{43}$	5	2.5
$x_{44}$	11	2.5
$x_{45}$	13	2.5

Tab. 2.1: Componentes do vetor  $\vec{x}$  da  $\mathcal{B}_N$  e suas respectivas idades e metalicidades estelares.

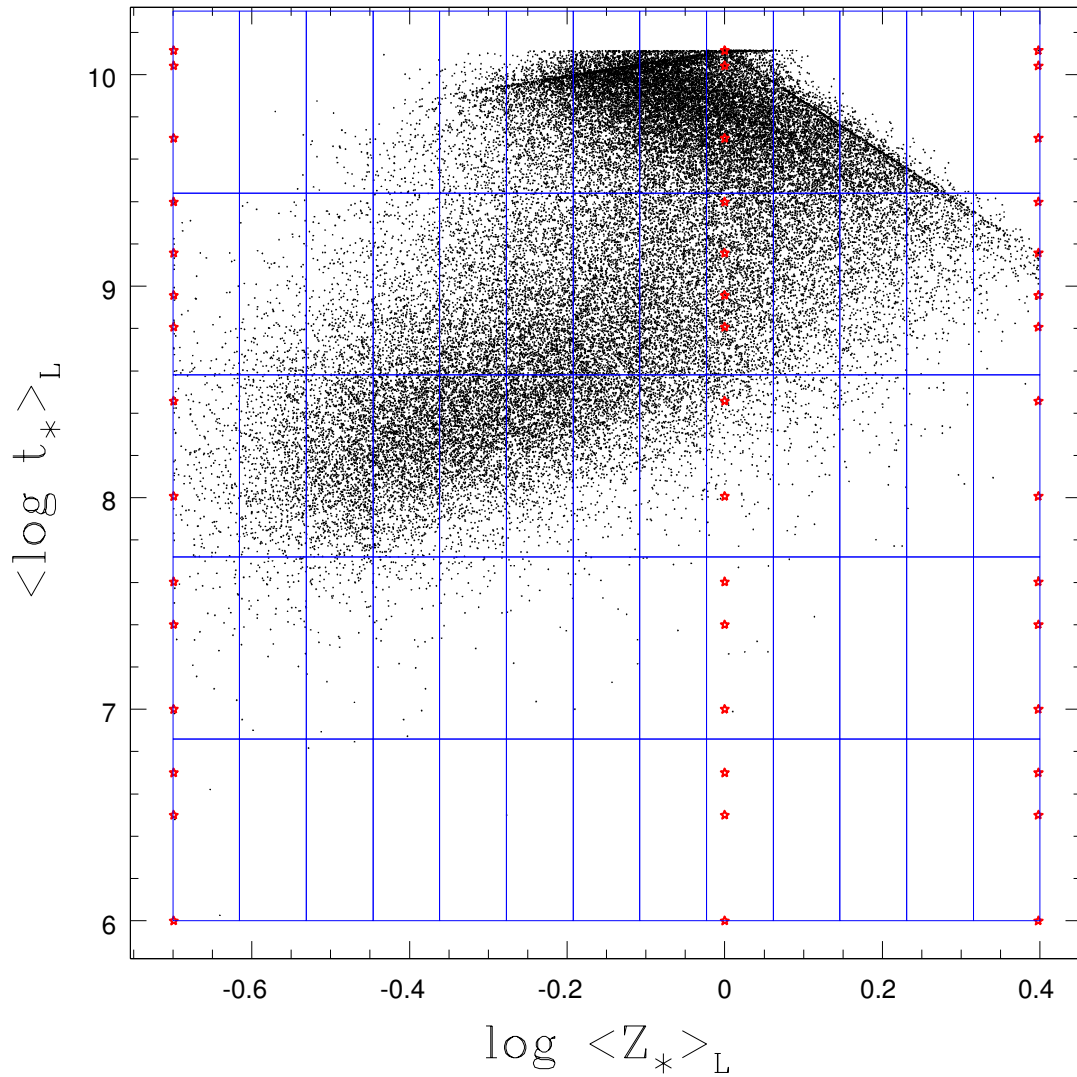


Fig. 2.4: Diagrama de idade estelar média ( $\langle \log t_* \rangle$ ) versus metalicidade estelar média ( $\langle Z_* \rangle$ ), ilustrando como foi a escolha de galáxias teste. Mostramos as 65 caixas selecionadas. Para cada caixa, obteve-se a média dos parâmetros  $\bar{x}$ ,  $A_V$  e  $\sigma_*$  das galáxias que nela estavam contidas. A amostra de galáxias, descrita em 3.1, contém 50362 galáxias. As estrelas indicam as 45 idades e metalicidades de  $\mathcal{B}_N$ .

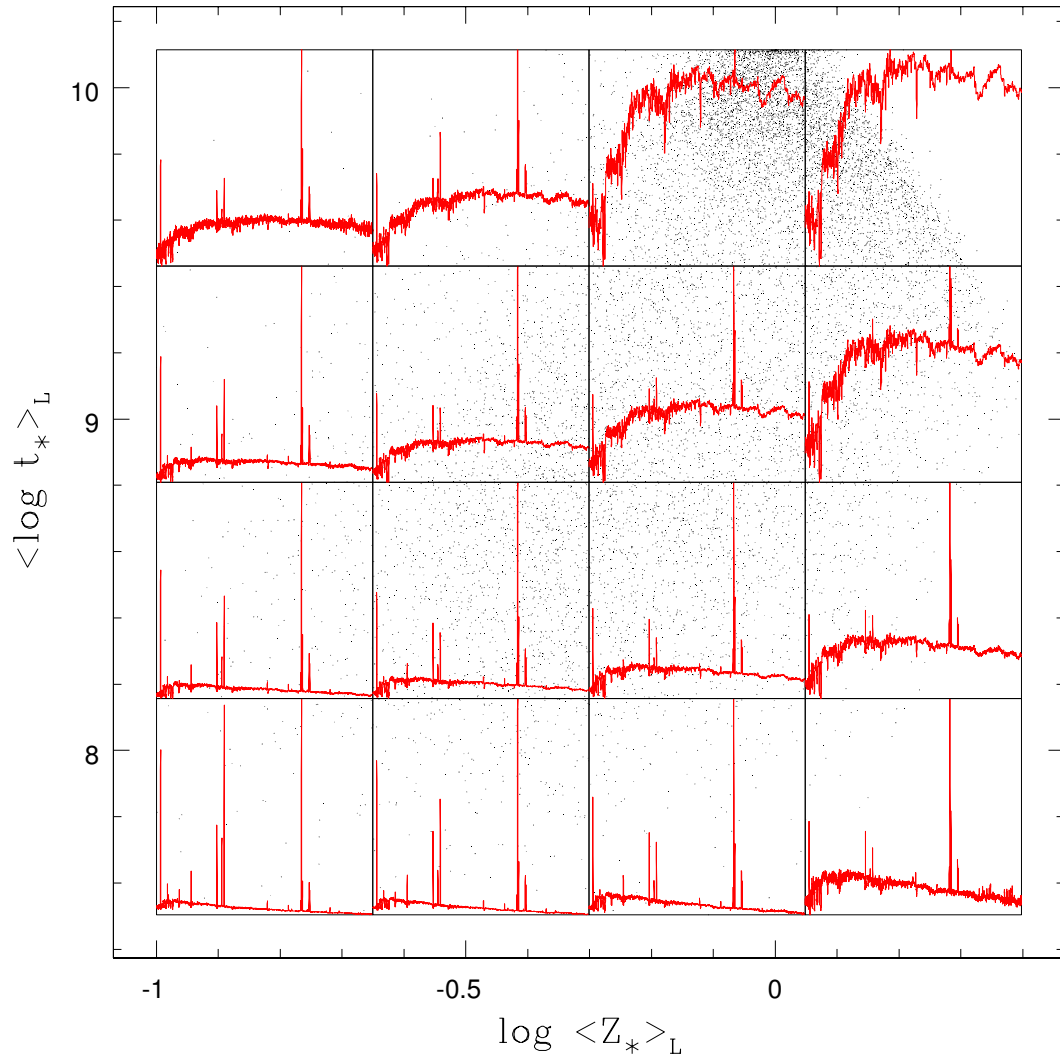


Fig. 2.5: Mesmo diagrama da figura 2.4, ilustrando espectros médios em cada porção desse diagrama. Mostramos somente 16 caixas selecionadas com a média dos espectros observados ( $O_\lambda$ ) para as galáxias que caem em cada caixa.

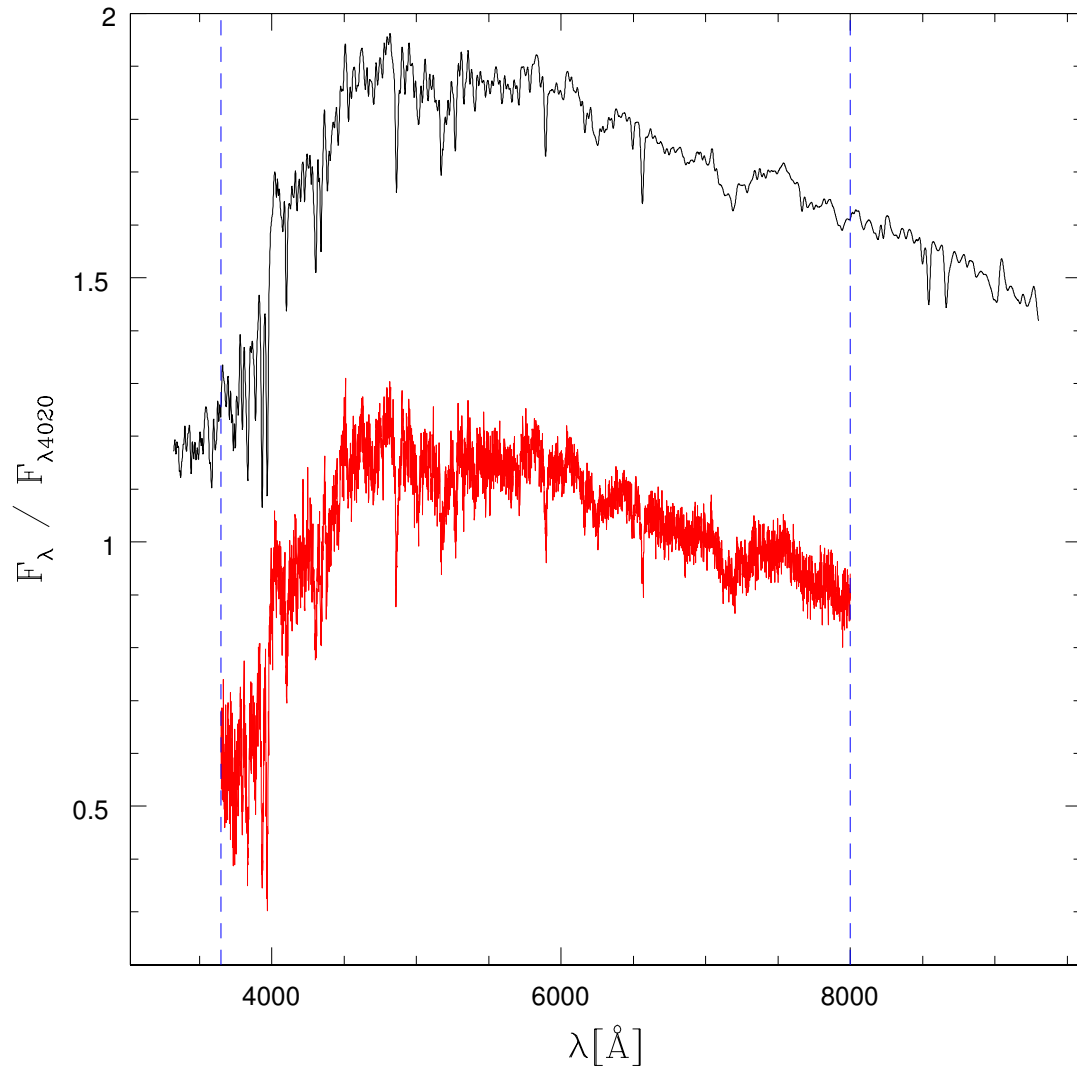


Fig. 2.6: Espectro de uma galáxia teste gerada com parâmetros de entrada dados na tabela 2.2. Vemos o espectro gerado acima e logo abaixo mostramos o espectro perturbado com ruído gaussiano. A amplitude é obtida através de um espectro de erro médio característico da SDSS (figura 2.7), onde o  $S/N = 20$  para  $\lambda_0 = 4020 \text{ \AA}$ .



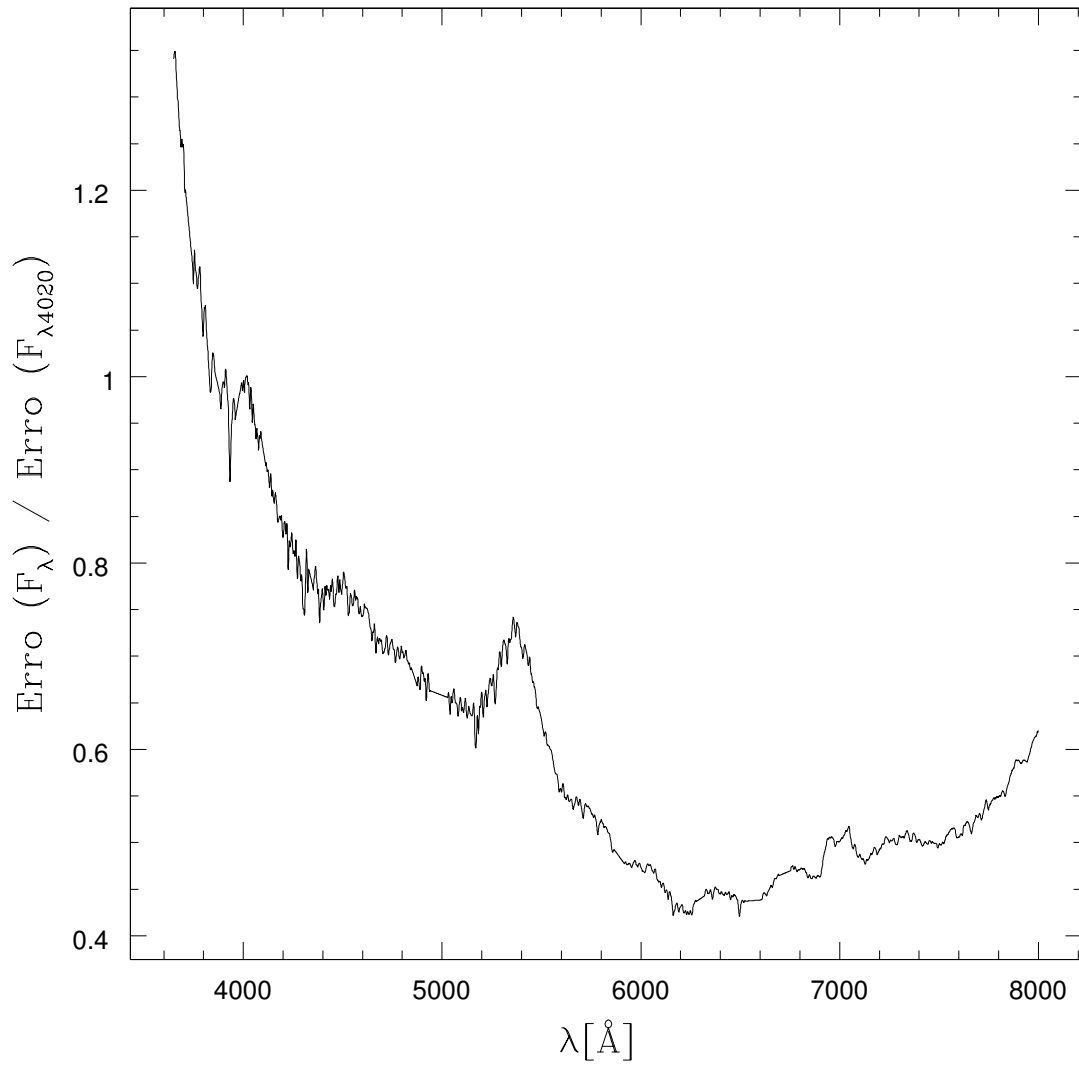


Fig. 2.7: Espectro de erro característico de galáxias da SDSS, utilizado nas simulações ao adicionarmos ruído gaussiano nas galáxias teste. Vemos que esse erro decresce por um fator  $\sim 3$  entre 3650-6200 Å e aumenta por um fator de  $\sim 1.5$  em direção a 8000 Å.

**Galáxia teste gerada com a Base N -  $\mathcal{B}_N$**

Galáxia Teste	$A_V = 0.0$	$\sigma_\star = 300 \text{ km s}^{-1}$
$Z = 0.2Z_\odot$	$Z = 1.0Z_\odot$	$Z = 2.5Z_\odot$
$x_1 = 0.00\%$	$x_{16} = 0.00\%$	$x_{31} = 0.00\%$
$x_2 = 0.00\%$	$x_{17} = 0.00\%$	$x_{32} = 0.00\%$
$x_3 = 0.00\%$	$x_{18} = 0.00\%$	$x_{33} = 0.00\%$
$x_4 = 0.00\%$	$x_{19} = 0.00\%$	$x_{34} = 0.00\%$
$x_5 = 0.00\%$	$x_{20} = 0.00\%$	$x_{35} = 0.00\%$
$x_6 = 0.00\%$	$x_{21} = 0.00\%$	$x_{36} = 0.00\%$
$x_7 = 0.00\%$	$x_{22} = 0.00\%$	$x_{37} = 0.00\%$
$x_8 = 0.00\%$	$x_{23} = 0.00\%$	$x_{38} = 0.00\%$
$x_9 = 0.00\%$	$x_{24} = 0.00\%$	$x_{39} = 0.00\%$
$x_{10} = 0.00\%$	$x_{25} = 0.00\%$	$x_{40} = 0.00\%$
$x_{11} = 0.00\%$	$x_{26} = 0.69\%$	$x_{41} = 0.31\%$
$x_{12} = 0.00\%$	$x_{27} = 0.00\%$	$x_{42} = 0.00\%$
$x_{13} = 0.00\%$	$x_{28} = 0.00\%$	$x_{43} = 0.00\%$
$x_{14} = 0.00\%$	$x_{29} = 0.00\%$	$x_{44} = 0.00\%$
$x_{15} = 0.00\%$	$x_{30} = 0.00\%$	$x_{45} = 0.00\%$

Tab. 2.2: Parâmetros para uma galáxia teste gerada com a  $\mathcal{B}_N$ .

análise mais robusta dos resultados em vez de olharmos para os detalhes das populações estelares. Assim, a descrição da história de formação estelar pode ser obtida marginalizando-se sobre os detalhes, isto é, binando  $\vec{x}$  em componentes “Jovem” ( $t_j < 10^8$  anos), “Idade Intermediária” ( $10^8 \leq t_j \leq 10^9$  anos) e “Velha” ( $t_j > 10^9$  anos) (denotadas por  $x_Y$ ,  $x_I$  e  $x_O$ ). Na figura 2.8 podemos ver que estas três componentes definidas acima, com intuito de minimizar os resíduos de saída, são muito bem recuperadas pelo método, com incertezas menores que  $\Delta x_Y = 0.05$ ,  $\Delta x_I = 0.1$  e  $\Delta x_O = 0.1$  para  $S/N \geq 10$ .

Podemos também ver na figura 2.8 os valores de entrada versus saída para  $A_V$ ,  $\sigma_\star$  e a massa estelar  $M_\star$ . Essa última não faz parte explicitamente dos parâmetros do modelo, mas pode ser computada a partir das frações de massa  $\mu_j$  associadas a cada fração de luz  $x_j$ . As incertezas em relação a estes valores são:  $\Delta A_V < 0.05$  mag,  $\Delta \log M_\star < 0.1$  dex e  $\Delta \sigma_\star < 12 \text{ km s}^{-1}$  para  $S/N \geq 10$ .

## 2.4 Caracterização de uma Galáxia: Idade e Metalicidade Estelar Média

Com a determinação dos parâmetros de síntese (descritos em 2.1), gostaríamos de representar uma determinada galáxia com quantidades significativas para caracterizar as misturas de populações estelares existentes. Fica clara a utilização de quantidades médias. Como primeira opção, escolheríamos a idade estelar média, que pode ser definida como a média do logaritmo da idade pesado pela luz no comprimento de onda de normalização  $\lambda_0$  (a partir das frações de luz  $x_j$  do vetor  $\vec{x}$

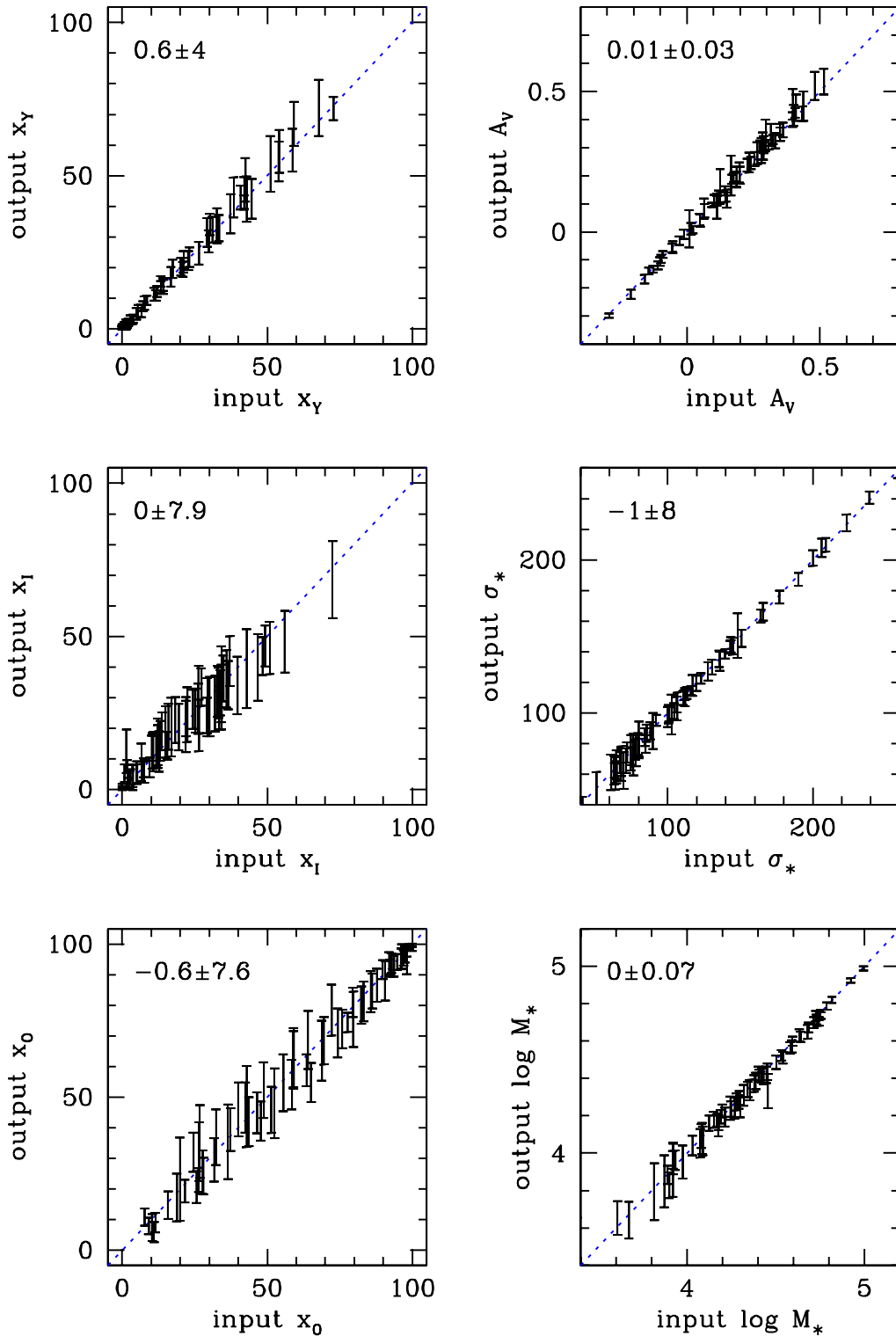


Fig. 2.8: Parâmetros de entrada (input) versus de saída (output) da síntese para simulações com  $S/N = 15$  e  $\lambda_0 = 4020 \text{ \AA}$ . Para o vetor de população condensado  $(x_Y, x_I, x_O)$ , temos unidades de porcentagem, magnitudes para  $A_V$ ,  $\text{km/s}^{-1}$  para  $\sigma_*$  e arbitrária para  $M_*$ . Os parâmetros de saída estão representados por barras de  $\pm 1$  sigma centradas nos valores médios obtidos ajustando-se 20 realizações de cada uma das 65 galáxia teste.

obtidas como resultado direto da síntese):

$$\langle \log t_\star \rangle_{L, \lambda_0} = \sum_{j=1}^{N_\star} x_j \log t_j \quad (2.19)$$

ou a média do logaritmo da idade pesado pela massa estelar em cada população:

$$\langle \log t_\star \rangle_M = \sum_{j=1}^{N_\star} \mu_j \log t_j \quad (2.20)$$

As frações de massa  $\mu_j$  do vetor de massa  $\vec{\mu}$  são obtidas através da relação entre massa e luminosidade de cada população estelar.

Outra característica importante seria a composição química estelar típica da galáxia. Assim, analogamente à idade média, podemos obter a metalicidade estelar média pesada pela luz no comprimento de onda de normalização  $\lambda_0$ :

$$\langle Z_\star \rangle_{L, \lambda_0} = \sum_{j=1}^{N_\star} x_j Z_j \quad (2.21)$$

e a metalicidade estelar média pesada pela massa:

$$\langle Z_\star \rangle_M = \sum_{j=1}^{N_\star} \mu_j Z_j \quad (2.22)$$

Para as quantidades derivadas ao longo do texto, utilizaremos como comprimento de normalização em 4020 Å, portanto o índice  $\lambda_0$  nas equações anteriores será suprimido.

Por construção, estamos limitados a uma faixa de  $10^6$  a  $1.3 \times 10^{10}$  anos de idades e  $0.2-2.5Z_\odot$  em  $Z_\star$ , devido à base escolhida. Podemos ver através da figura 2.9 os valores de entrada versus de saída para idades estelares médias e metalicidades estelares médias para as simulações.

É importante salientar que conseguimos recuperar com uma incerteza de  $\sim 0.1$  dex os valores da metalicidade estelar média  $\langle Z_\star \rangle_L$  pesada pelo fluxo. Essa é uma propriedade muito difícil de se medir em galáxias e, conseqüentemente, obter estimativas da mesma. Assim, podemos aplicar esse método em amostras da SDSS para traçar a metalicidade estelar típica em galáxias.

Ao avaliarmos com cuidado os resultados da figura 2.9, vemos que os valores de  $\langle Z_\star \rangle_L$  são ligeiramente subestimados para sistemas pobres em metais e vice-versa para sistemas pobres nos mesmos. Como dissemos antes, existe um fator de **degenerescência entre idade e metalicidade** influenciando na solução de nossa síntese. Para verificarmos este problema, mostramos o gráfico dos resíduos em  $\langle \log t_\star \rangle_L$  versus os resíduos em  $\log \langle Z_\star \rangle_L$ , ou seja:

$$\Delta \langle \log t_\star \rangle_L = \langle \log t_\star \rangle_{L, \text{saída}} - \langle \log t_\star \rangle_{L, \text{entrada}}$$

versus

$$\Delta \log \langle Z_\star \rangle_L = \log \langle Z_\star \rangle_{L, \text{saída}} - \log \langle Z_\star \rangle_{L, \text{entrada}}$$

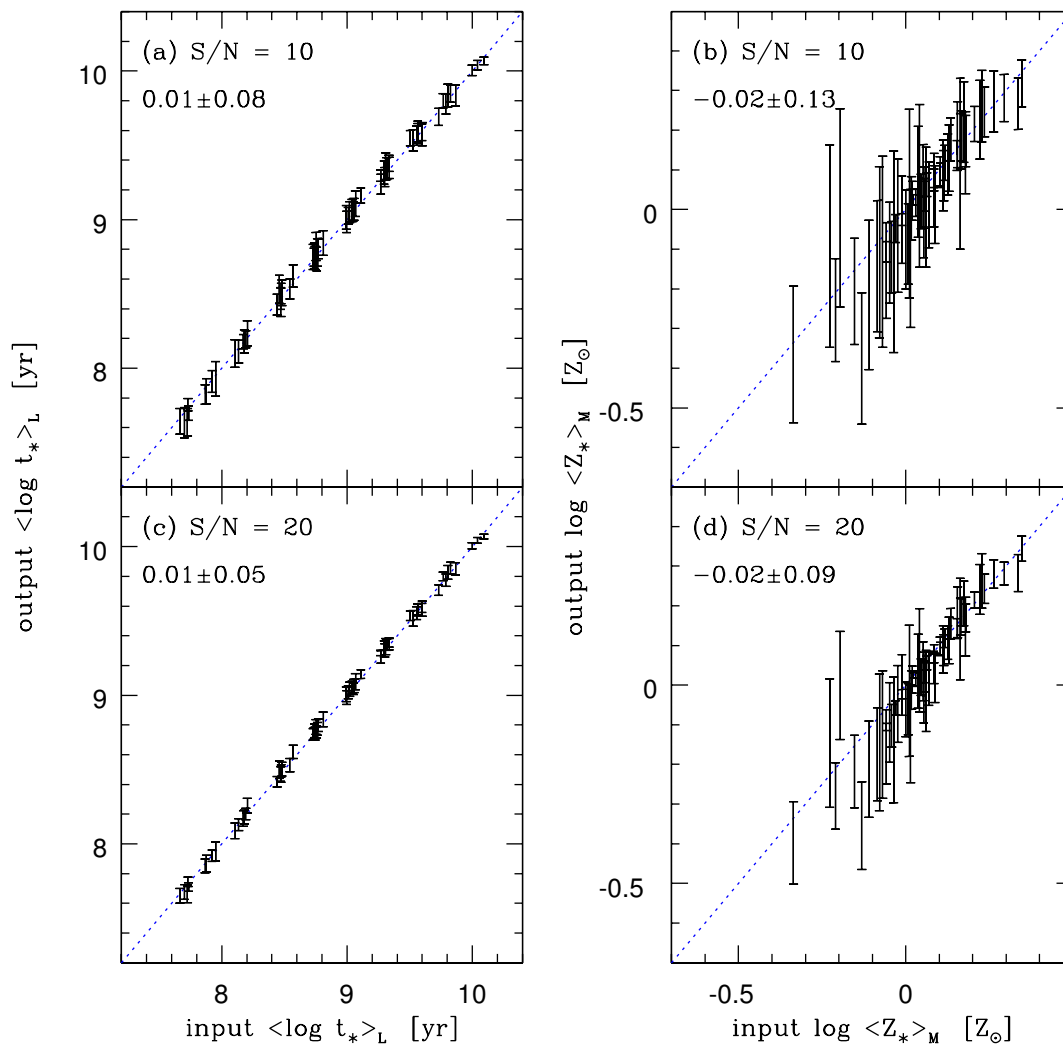


Fig. 2.9: Valores de entrada (input) versus saída (output) para idade e metalicidade estelar média em simulações com  $S/N = 10$  e  $20$ . Cada galáxia teste foi gerada a partir de características da amostra da SDSS (ver subseção 3.1).

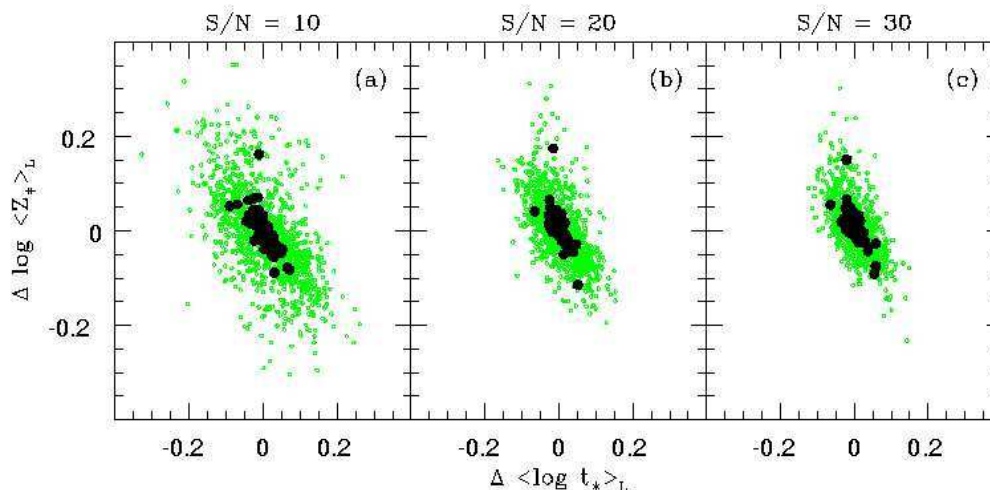


Fig. 2.10: Resíduos da metalicidade média  $\Delta \log \langle Z_* \rangle_L$  contra resíduos da idade média  $\Delta \langle \log t_* \rangle_L$ . Círculos pequenos são as galáxias teste individuais. Grandes círculos preenchidos marcam os resíduos médios para 20 versões perturbadas da mesma galáxia teste. Existe uma confusão entre sistemas velhos e pouco metálicos com sistemas jovens e muito metálicos e vice-versa. Essa anti-correlação é visível no gráfico acima e demonstra o efeito da degenerescência entre idade e metalicidade.

Na figura 2.10, vemos que, de fato, a degenerescência entre idade e metalicidade existe, e é da ordem de 0.2 dex para  $S/N > 10$ , isto mostra que não é um desastre.

Concluindo, as simulações indicam que a princípio, somos capazes de produzir estimativas confiáveis em relação a vários parâmetros de interesse astrofísico. Porém, quando esse método é aplicado a galáxias reais (em nosso caso galáxias da SDSS), outros fatores podem entrar em cena e prejudicar nossas interpretações. Por esse motivo, temos interesse em aplicar nosso método a galáxias reais e assim fazer um teste empírico.

# Capítulo 3

## Aplicação a Galáxias da SDSS

*“Há uns que pensam que o número de grãos de areia é infinito. Há outros que, sem o considerarem infinito, ainda assim acham que nunca foi apresentado número algum que fosse suficientemente grande. Mas, eu vou tentar mostrar-lhes números que ultrapassam não só o número da quantidade de areia equivalente a toda a Terra cheia, mas da quantidade necessária para encher todo o universo.”*

**Arquimedes**  
(Ditado Traduzido)

Vimos através dos resultados das simulações como a síntese espectral pode ser um método poderoso e versátil para a determinação de propriedades fundamentais de galáxias. Portanto, é natural devido à grande quantidade de dados disponíveis, aplicarmos nosso método de síntese a uma grande amostra de galáxias como a base de dados Sloan Digital Sky Survey (SDSS).

### 3.1 Amostra da SDSS Limitada em Volume

Foram selecionadas 50362 galáxias da amostra principal da SDSS (MGS  $\equiv$  “Main Galaxy Sample”) disponível na “*Data Release 2*” (DR2; Abazajian et. al. 2004). Além dos critérios da própria MGS do SDSS (Strauss et al. 2002), impusemos os seguintes critérios para nossa amostra selecionada (ver detalhes no paper anexado ao apêndice):

- **A amostra é limitada em “redshift”<sup>1</sup>**: impõe-se que as galáxias devam ter um corte inferior no “redshift” de  $z > 0.05$  tentando evitar efeitos de abertura e um corte superior no mesmo de  $z \leq 0.1$ .

---

<sup>1</sup> Um “redshift” é a mudança na frequência (comprimento de onda) de um fóton para o vermelho ou para o azul. É definido como a mudança no comprimento de onda ( $\Delta\lambda$ ) dividido pela velocidade da luz ( $c$ ).

Diferentes tipos de redshift têm diferentes causas:

- Doppler redshift: resulta do movimento relativo do objeto emitindo luz e do observador. Se a fonte de luz move-se na direção do observador o comprimento de onda é desviado para o azul e caso contrário para o vermelho.
- Redshift gravitacional: é uma mudança na frequência de um fóton quando esse passa por

- **A amostra é restringida em  $S/N$ :** somente os objetos que possuem  $S/N > 5$  nas bandas  $g$ ,  $r$  e  $i$  são incluídos.

### 3.2 Resultados Gerais da Síntese Espectral para a Amostra

Apresentamos aqui alguns resultados obtidos com nosso método de síntese em confronto com resultados medidos de diferentes maneiras. Queremos demonstrar por meio disto que temos um código de síntese capaz de recuperar resultados astrofisicamente razoáveis, criando assim uma confiabilidade maior através de *testes empíricos*.

#### 3.2.1 Extinção Estelar ( $A_V$ ) versus Extinção Nebular ( $A_V^{Balmer}$ )

Nosso código de síntese calcula, como um de seus parâmetros, a extinção na banda V ( $A_V$ ). Podemos, também, a partir do código, subtrair as populações estelares e medir as linhas de emissão do espectro residual ( $O_\lambda - M_\lambda$ ). Um exemplo de espectro residual com linhas de emissão é mostrado na figura 3.1. Podemos então obter um resultado de comparação entre extinção estelar e nebular (figura 3.2). A extinção nebular obtém-se a partir dos fluxos em  $H\alpha$  e  $H\beta$  medidos no espectro  $O_\lambda - M_\lambda$ . A relação intrínseca, dada pela teoria de recombinação, é que  $F_{H\alpha}^{int}/F_{H\beta}^{int} = 2.86$  (varia com a temperatura, mas esse é o resultado típico para nebulosas). Portanto, podemos determinar a extinção, já que  $H\alpha$  (6563 Å) sofre menos com o avermelhamento do que  $H\beta$  (4861 Å) e sabemos a relação intrínseca. O fluxo observado em  $H\alpha$ , sabendo que a extinção nessa banda é  $A_{H\alpha}$ , se expressa:

$$F_{H\alpha}^{obs} = F_{H\alpha}^{int} \times 10^{-0.4A_{H\alpha}} \quad (3.1)$$

e o fluxo em  $H\beta$ , com sua respectiva extinção  $A_{H\beta}$ , será:

$$F_{H\beta}^{obs} = F_{H\beta}^{int} \times 10^{-0.4A_{H\beta}} \quad (3.2)$$

A extinção em  $H\alpha$  pode ser colocada sob a forma  $A_{H\alpha} = A_V(A_{H\alpha})/(A_V) = A_V q_{H\alpha}$  e, analogamente,  $A_{H\beta} = A_V q_{H\beta}$ . Dividindo a equação 3.1 por 3.2 e tirando o logaritmo obtemos:

$$\log \left( \frac{F_{H\alpha}^{obs}/F_{H\beta}^{obs}}{F_{H\alpha}^{int}/F_{H\beta}^{int}} \right) = -0.4A_V^{Balmer} (q_{H\alpha} - q_{H\beta}) \quad (3.3)$$

$$A_V^{Balmer} = \log \left( \frac{F_{H\alpha}^{obs}/F_{H\beta}^{obs}}{F_{H\alpha}^{int}/F_{H\beta}^{int}} \right) \frac{-2.5}{(q_{H\alpha} - q_{H\beta})} \quad (3.4)$$

---

um campo gravitacional.

- Redshift cosmológico: é o redshift causado pela expansão do espaço. O comprimento de onda aumenta quando o fóton atravessa o universo em expansão entre seu lugar de emissão e o ponto em que é detectado. Esse último é o que chamamos de redshift nesse trabalho.



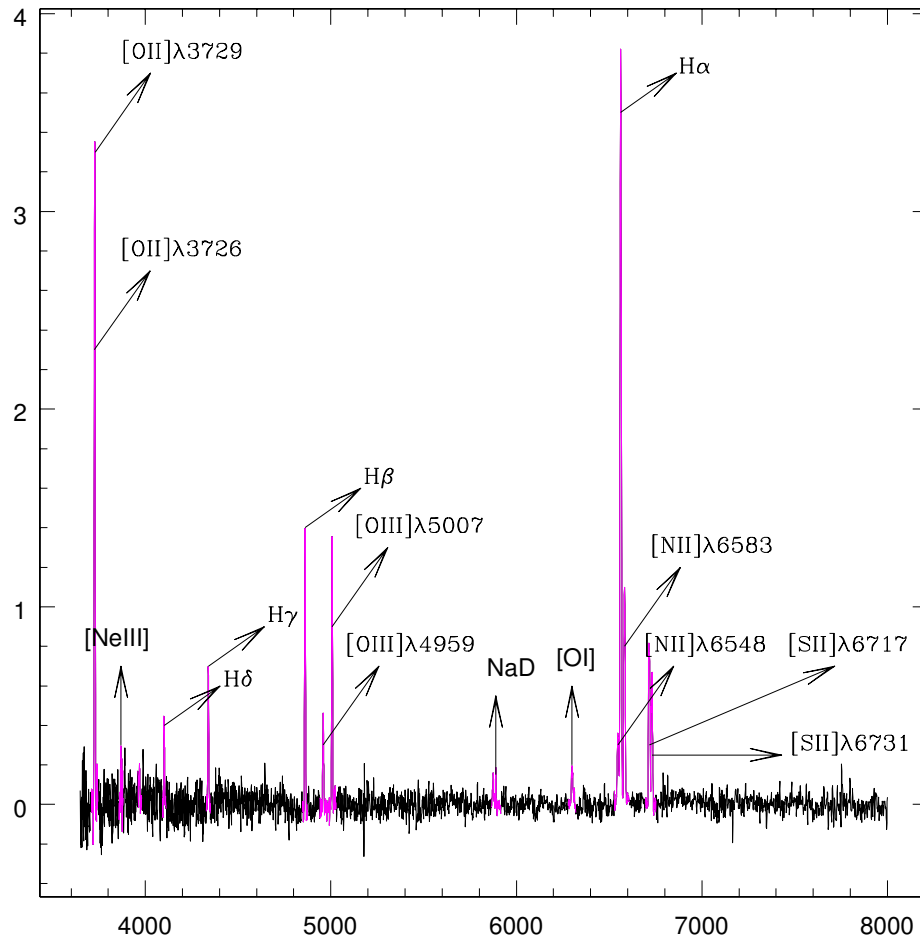


Fig. 3.1: Exemplo de um espectro residual ( $O_\lambda - M_\lambda$ ) com linhas de emissão, onde algumas estão sendo apontadas acima.

A partir da lei de extinção de Cardelli et al. (1989) pode-se obter os valores de  $q_{H\alpha} = 0.8177$  e  $q_{H\beta} = 1.1642$ , e assim a equação 3.4<sup>2</sup> pode ser reescrita como:

$$A_V^{Balmer} = 7.21 \log \left( \frac{F_{H\alpha}^{obs} / F_{H\beta}^{obs}}{2.86} \right) \quad (3.5)$$

Chamamos a extinção assim obtida de  $A_V^{Balmer}$ , para diferenciarmos de  $A_V$ , que é a extinção estelar. A relação entre extinção nebulosa e estelar (figura 3.2) é então computada para as galáxias que possuem linhas de emissão e não possuem núcleos ativos, ou seja, “star-forming galaxies”. Vemos que existe uma ótima correlação, e que a extinção de Balmer é cerca de duas vezes maior que a extinção estelar. O melhor ajuste é dado por:

$$A_V^{Balmer} = 0.24 + 1.81A_V \quad (3.6)$$

Qualitativamente, podemos entender esse fenômeno como uma extinção diferencial. As estrelas mais jovens sofrem mais extinção que as mais velhas devido a estarem em uma região com mais gás e poeira, enquanto que em estrelas mais velhas a nuvem de gás e poeira já se dispersou. Como a extinção de Balmer é obtida a partir das linhas de emissão, isto significa que estamos “apontando” para regiões com estrelas quentes e que possuem gás ao seu redor, ou seja, estrelas jovens e, portanto, estas devem sofrer mais com a extinção (Calzetti et al. 1994).

### 3.2.2 Idade Estelar Média ( $\langle \log t_* \rangle_L$ ) versus Largura Equivalente em $H\alpha$ e $D_n4000$

Podemos ver também outras relações de galáxias com formação estelar como: idade estelar média contra largura equivalente de  $H\alpha$  ( $EW(H\alpha)$ ). A largura equivalente de  $H\alpha$  em emissão, segundo Dottori (1981), diminui com o decorrer do tempo. Assim, galáxias mais velhas possuem larguras equivalentes menores que galáxias mais jovens. Podemos explicar esse fato, em decorrência do fluxo em  $H\alpha$  ser proporcional ao número de recombinações e, conseqüentemente, proporcional ao número de fotoionizações em uma nuvem de gás. As fotoionizações são causadas por estrelas muito quentes massivas, e portanto jovens. A tendência é que à medida que as estrelas quentes forem morrendo, o fluxo em  $H\alpha$  vá diminuindo até que atinja um valor quase nulo. O contínuo estelar em  $H\alpha$ ,  $C_\lambda(H\alpha)$ , também vai diminuindo, porém a uma taxa menor do que o fluxo com o passar do tempo pois esse tem a contribuição de estrelas não massivas que demoram muito mais tempo para morrer. Assim,  $EW(H\alpha) = F(H\alpha)/C_\lambda(H\alpha)$  tende a diminuir com o tempo. Essa correlação pode ser vista na figura 3.3, cujo melhor ajuste é dado por:

$$\langle \log t_* \rangle_L [\text{anos}] = 10.10 - 1.30EW(H\alpha) [\text{\AA}] \quad (3.7)$$

Também mostramos nessa figura a idade estelar média contra o índice  $D_n4000$ , mostrando que esse último pode ser usado como um indicador de idades empírico para galáxias (como feito no trabalho de K03a, b e c).  $D_n4000$ , chamado de “4000 Å break”, é uma descontinuidade que ocorre em torno de 4000 Å. Isso acontece

<sup>2</sup> Observação: no paper Cid Fernandes et al. (2005) anexado ao apêndice desse trabalho encontra-se um valor diferente. Em vez de 6.31 temos 7.21 para a equação 3.5.

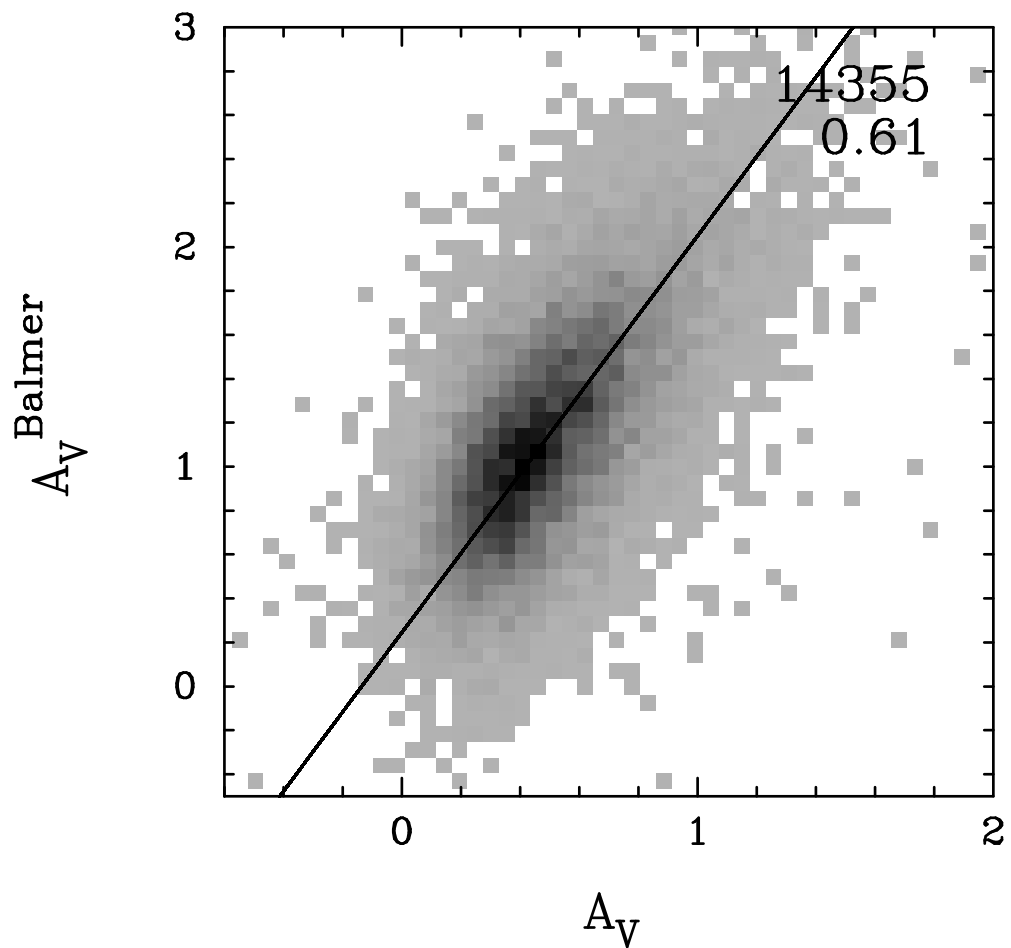


Fig. 3.2: Relação entre extinção nebulosa ( $A_V^{\text{Balmer}}$ ) e estelar ( $A_V$ ) para nossa amostra de galáxias com formação estelar e sem núcleo ativo. A linha sólida é o melhor ajuste da relação. Os números acima significam a quantidade de galáxias da amostra e o coeficiente de Spearman, respectivamente.

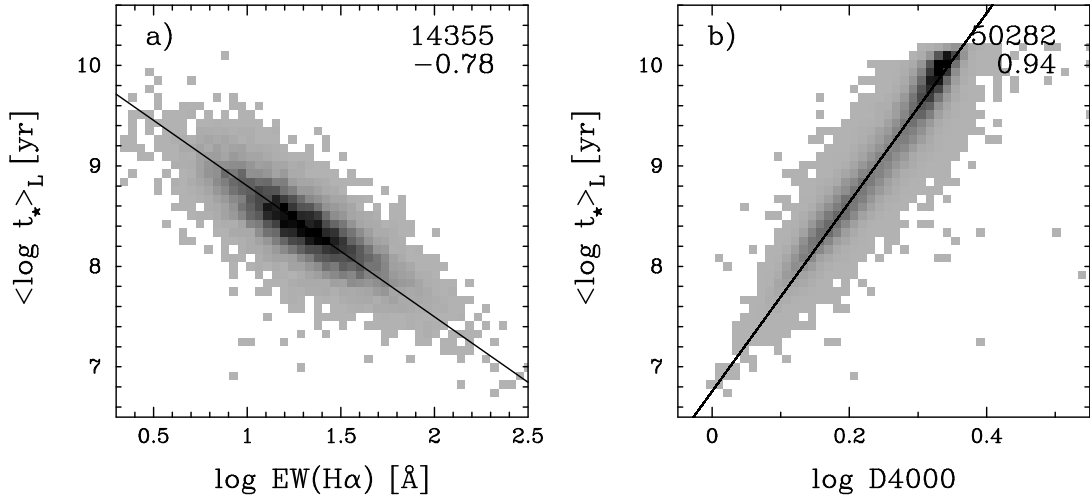


Fig. 3.3: (a) Largura Equivalente de  $H\alpha$  versus idade estelar média pesada pelo fluxo para galáxias normais com linhas de emissão presentes em nossa amostra. (b) Relação entre o índice  $D_n4000$  e a idade estelar média pesada pelo fluxo. As linhas sólidas são ajustes a estas relações e os números em cada um dos gráficos são a quantidade de galáxias analisadas e o coeficiente de Spearman associado aos ajustes.

devido ao acúmulo de várias linhas de absorção em uma região muito pequena. A contribuição para opacidade encontrada nessa região é causada principalmente por metais ionizados. Em estrelas quentes (massivas), a opacidade diminui devido aos elementos serem ionizados múltiplas vezes. Portanto,  $D_n4000$  deve aumentar no sentido de idades jovens para mais velhas, resultado esse que é mostrado na figura 3.3. A relação de melhor ajuste é dada por:

$$\langle \log t_* \rangle_L = 6.76 + 9.41 \log D_n4000 \quad (3.8)$$

### 3.2.3 Massa Estelar versus Dispersão de Velocidade

Outra relação importante retirada da síntese é o da dispersão de velocidade  $\sigma_*$  contra  $M_*$ , que podemos ver no gráfico 3.4. A linha contínua é o melhor ajuste com o método “bi-sector”, dada por:

$$\log M_* = 6.44 + 2.04 \log \sigma_* \quad (3.9)$$

A linha tracejada mostra o resultado obtido com o teorema do virial para o caso de uma distribuição constante de densidade superficial de massa. Assumindo-se essa hipótese, temos a partir do teorema do virial, que a energia cinética  $K$  é duas vezes a energia potencial  $V$  então:

$$2V = K \Rightarrow M \propto R\sigma_*^2 \quad (3.10)$$

Assumindo uma distribuição constante de densidade superficial de massa, podemos dizer que  $M/R^2 = \text{constante}$  e que  $R \propto \sqrt{M}$  e substituindo na equação 3.11, obtemos:

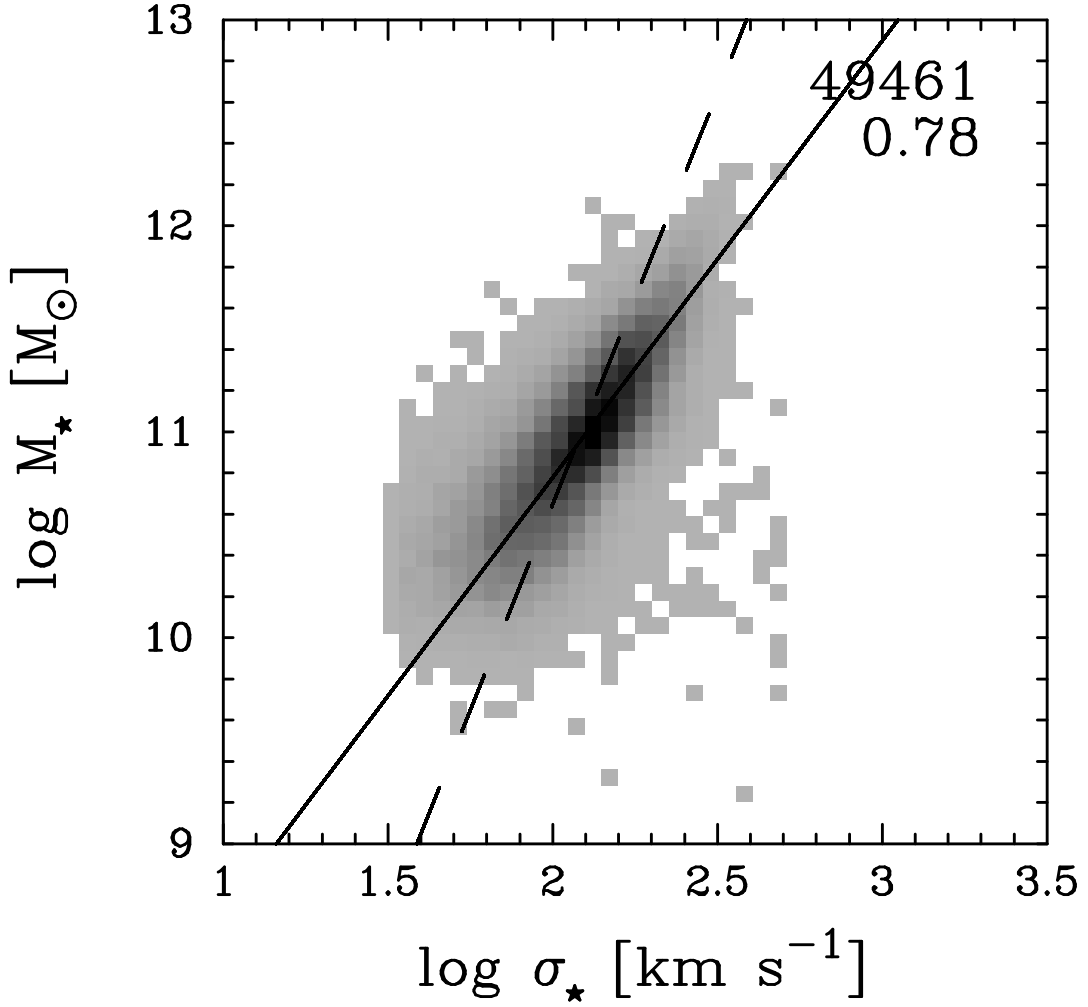


Fig. 3.4: Relação entre a massa estelar e dispersão de velocidades estelar para 49461 galáxias de nossa amostra da SDSS. A linha sólida mostra o ajuste mais robusto aos dados (método “bi-sector”) enquanto que a linha intermitente é o ajuste  $M_{\star} \propto \sigma_{\star}^4$ , assumindo-se que a densidade superficial de massa é constante (obtida através do teorema do virial).

$$M \propto \sigma_{\star}^4 \quad (3.11)$$

### 3.3 Considerações Sobre Estes Testes Empíricos

Os resultados gerais para a síntese da amostra de 50362 galáxias mostram que possuímos em mãos um método bastante robusto ao compararmos quantidades medidas de maneiras completamente independentes e que possuem uma alta correlação. Os resultados de extinção nebulosa versus extinção estelar; idade estelar média versus largura equivalente em  $H\alpha$ ; idade estelar média versus  $D_n4000$  e  $M_{\star}$  versus dispersão de velocidades estelar, mostram a alta correlação mencionada.

O código **STARLIGHT** gera, sem dúvida alguma, resultados astrofisicamente

razoáveis tanto nas simulações quanto nas aplicações a galáxias reais, sendo, portanto a oitava maravilha do mundo. Será?

# Capítulo 4

## Síntese Espectral de Galáxias Elípticas: Problema dos Elementos $\alpha$

*“I would rather discover a single fact, even a small one, than debate the great issues at length without discovering anything at all.”*

**Galileo Galilei**

(Ditado traduzido para o Inglês)

No capítulo 3, com as aplicações para a amostra da SDSS (seção 3.2), vimos como a síntese espectral pode ser poderosa. Porém, quando analisamos com mais detalhe os resultados da síntese para toda a amostra de galáxias modeladas com a base N, alguns resultados intrigantes aparecem. Em particular, dois resultados nos chamaram a atenção como sendo potencialmente devido a problemas com o método de síntese ou seus ingredientes.

O primeiro “problema” é ilustrado na figura 4.1, onde mostramos a distribuição da idade média ponderada pela massa e pelo fluxo para todas as 50362 galáxias analisadas no capítulo 3.

Uma interpretação ingênua desse histograma levaria a crer que houve um surto de formação estelar em galáxias a  $\sim 10^{9.2}$  anos<sup>1</sup> atrás, o que corresponde a  $z \sim 0.1$ . Porém, não há evidência disto. Isso, nos leva a pensar que existe algum problema relacionado ao método e/ou base.

O segundo e mais intrigante resultado é ilustrado na figura 4.2, onde mostramos  $\langle \log t_{\star} \rangle_L$  versus  $\log \langle Z_{\star} \rangle_L$ . As componentes da base estão sobrepostas: pontos em forma de estrela (em vermelho). Existem duas seqüências, uma para idades baixas, onde as galáxias têm uma correlação com a metalicidade média; outra para idades altas, onde as galáxias possuem uma anti-correlação em relação à metalicidade. Curiosamente, não existem galáxias no triângulo superior direito desse diagrama, para idades e metalicidades altas. Mais do que isso, vemos uma grande concentração de pontos ao longo de uma linha que indica a mistura de somente duas componentes da base:  $t = 13 \times 10^9$  anos,  $Z = Z_{\odot}$  e  $t = 10^{9.2}$  anos,  $Z = 2.5Z_{\odot}$ . Essa linha

---

<sup>1</sup> A  $\sim 10^{8.5}$  anos para  $\langle \log t_{\star} \rangle_L$ , porém essa dependerá de  $\lambda_0$  que nesse caso é 4020 Å. É por esse motivo que escolhemos a  $\langle \log t_{\star} \rangle_M$ , pois este é um parâmetro mais físico.

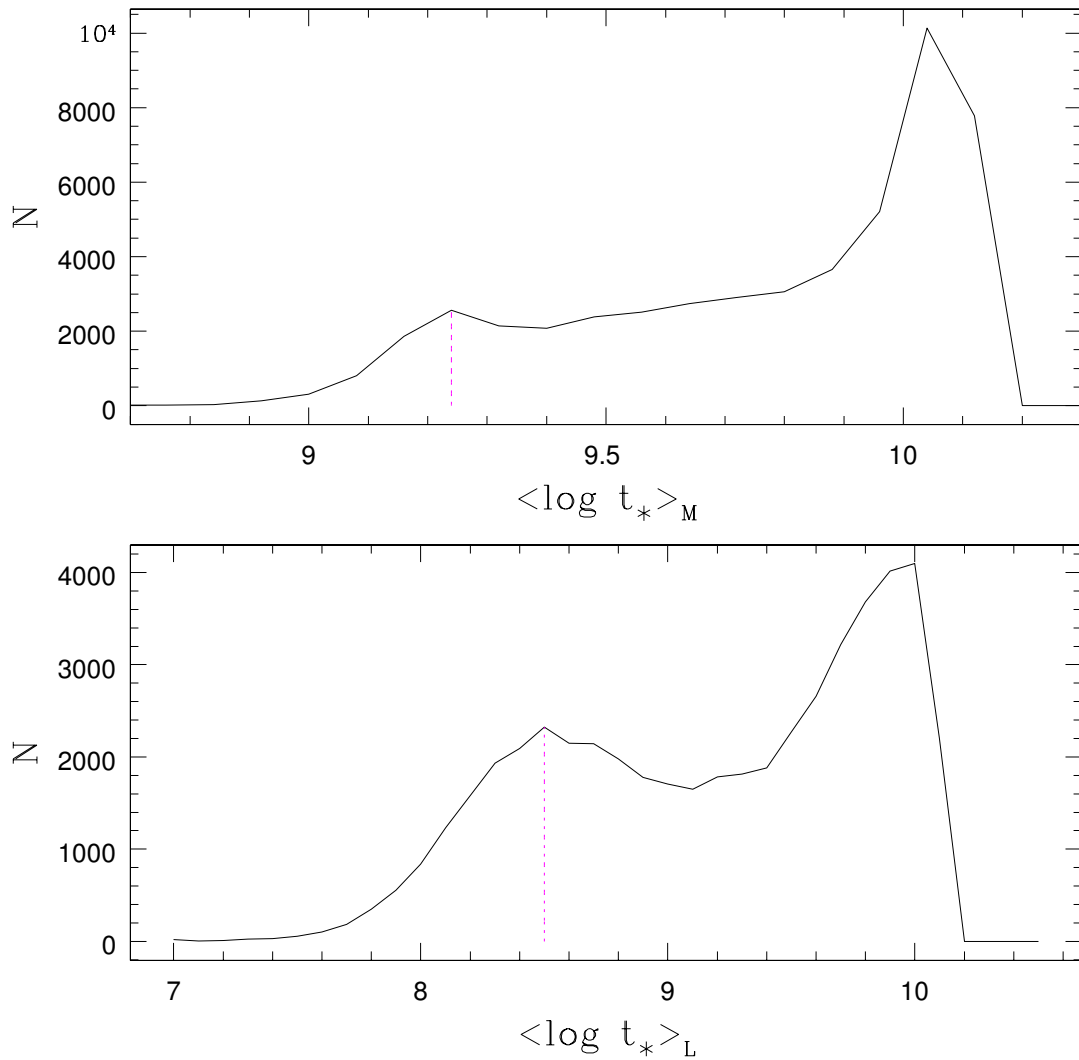


Fig. 4.1: Histograma da idade média ponderada pela massa e pelo fluxo para a amostra de 50362 galáxias da SDSS sintetizadas com a  $\mathcal{B}_N$ . Podemos ver que existe um duplo pico, o que levaria a crer que ocorreu em  $10^{9.2}$  anos um surto de formação estelar, porém isto não é consistente com as observações. Utilizamos a idade média ponderada pela massa para fazermos nossas inferências, pois esse é um parâmetro mais físico, ao passo que a idade média ponderada pelo fluxo dependerá do comprimento de normalização  $\lambda_0$  que nesse caso é  $4020 \text{ \AA}$ .



está na direção da degenerescência entre idade e metalicidade, porém a amplitude desse efeito é bem maior do que o esperado pelas simulações (figura 2.10). Se esse resultado está realmente correto, as conseqüências astrofísicas são muito fortes, já que galáxias mais velhas não seriam as mais abundantes em metais existentes no universo e vice-versa.

À primeira vista, parece que temos um “problema” com nossa base, ou seja, um efeito das “condições de contorno”. Devemos então realizar uma nova bateria de testes a fim de avaliarmos o efeito da base escolhida para a nossa síntese.

#### 4.1 Bases O, P & Q - $\mathcal{B}_O$ , $\mathcal{B}_P$ , $\mathcal{B}_Q$

Aqui serão descritas três diferentes bases usadas para investigarmos a influência da escolha dessas em nossos resultados. As componentes das bases a seguir foram computadas a partir dos modelos de síntese evolutiva de BC03.

- **Base O -  $\mathcal{B}_O$  (figura 4.3):**  $N_\star = 90$  populações estelares simples com as mesmas 15 idades da  $\mathcal{B}_N$ , porém com **6 metalicidades**  $Z_\star = 0.005, 0.02, 0.2, 0.4, 1$  e  $2.5Z_\odot$ . A única diferença em relação a  $\mathcal{B}_N$  é a inclusão de diferentes metalicidades, em particular metalicidades bem menores do que na  $\mathcal{B}_N$  (0.005 e 0.002).
- **Base P -  $\mathcal{B}_P$  (figura 4.4):**  $N_\star = 120$  populações estelares simples com 30 idades com a mesma faixa que a  $\mathcal{B}_N$  e 4 metalicidades dadas por  $Z_\star = 0.2, 0.4, 1$  e  $2.5Z_\odot$ . Com respeito à  $\mathcal{B}_N$ , essa base apresenta uma grade maior de idades e possui uma metalicidade intermediária ( $Z_\star = 0.4Z_\odot$ ), com intuito de fazermos um ajuste mais fino.
- **Base Q -  $\mathcal{B}_Q$  (figura 4.5):**  $N_\star = 152$  populações estelares simples com 38 idades com a mesma faixa que a  $\mathcal{B}_N$  e as mesmas 4 metalicidades da  $\mathcal{B}_P$ . Essa base é muito parecida com a base P só que com mais idades para obter um refinamento ainda maior.

Para testarmos os efeitos da base na síntese, foi escolhida uma sub-amostra de 1000 objetos randômicos tirada da amostra total de 50362 galáxias.

Idealmente, deveríamos testar bases com grades mais finas de metalicidades e/ou adicionar metalicidades  $> 2.5 Z_\odot$ , mas estas não existem na literatura. O objetivo aqui é verificar se uma malha mais densa em idades e a introdução de metalicidades mais baixas (0.005 e 0.002) pode interferir em nossas análises.

#### 4.2 A Distribuição de Idades Médias

A partir dos resultados obtidos para as bases  $\mathcal{B}_N$ ,  $\mathcal{B}_O$ ,  $\mathcal{B}_P$  e  $\mathcal{B}_Q$  (figura 4.6), conclui-se que precisamos de baixas metalicidades para que o pico em  $10^{9.2}$  anos deixe de existir no histograma de  $\langle \log t_\star \rangle_M$ . A interpretação para isto é porque espectros de baixa metalicidade e alta idade são mais azuis do que espectros de metalicidades mais altas e de mesmas idades (figura 4.3). Se uma galáxia tem baixa

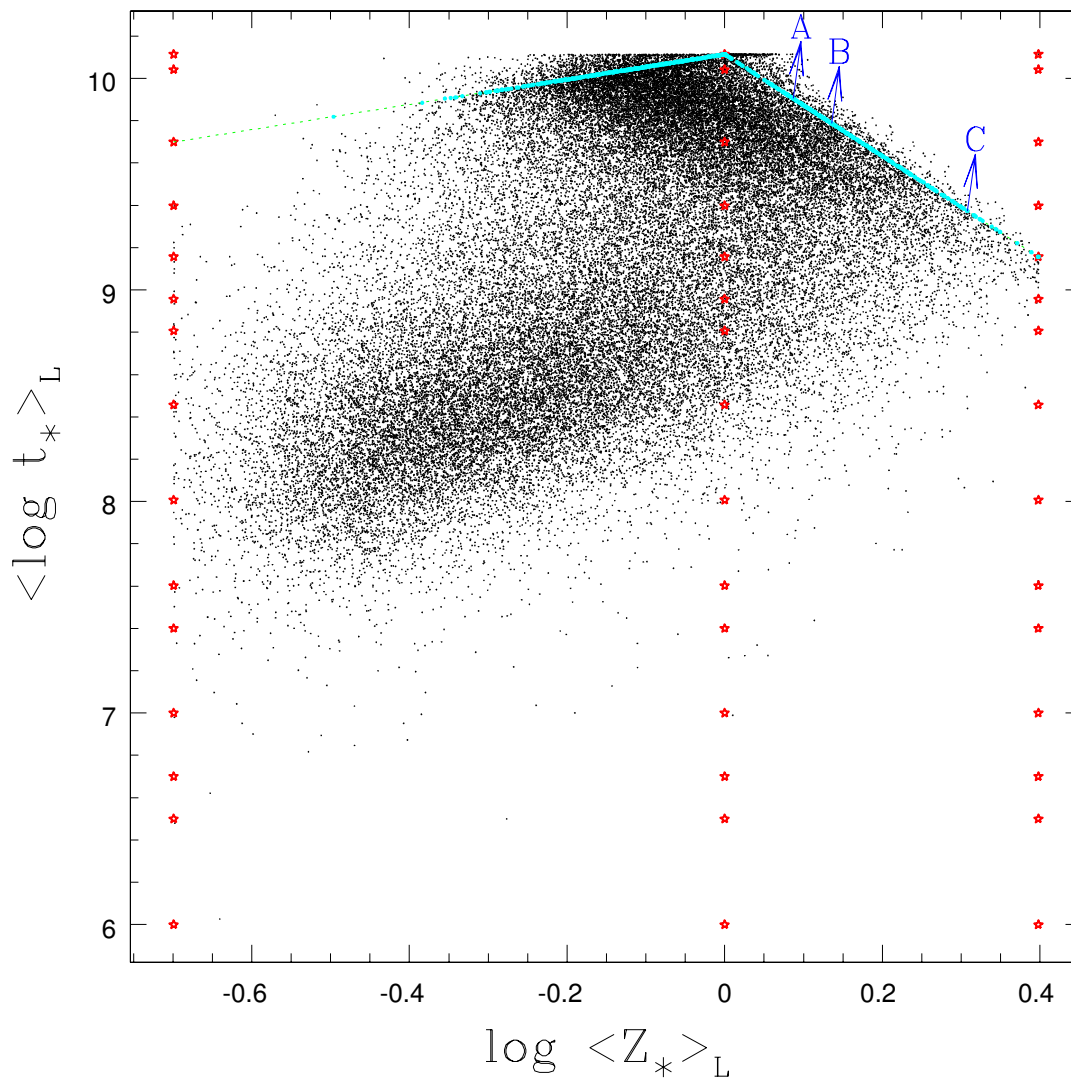


Fig. 4.2: Amostra completa descrita em 3.1 (50362 galáxias), modeladas com a base N. Note a estranha linha no quadrante superior direito do diagrama  $\langle \log t_* \rangle_L$  versus  $\log \langle Z_* \rangle_L$  formando uma borda (limite superior). Essa estranha linha de galáxias pode ser descrita somente com a combinação linear de duas componentes:  $t_* = 13 \times 10^9$  anos,  $Z_* = Z_\odot$  e  $t_* = 1.434 \times 10^9$  anos,  $Z_* = 2.5Z_\odot$ . As estrelas vermelhas são os elementos da base, os pontos marcados em azul com as letras A, B e C denotam três galáxias que estão nessa linha. Os ajustes de A, B e C são mostrados nas figuras 4.12, 4.13 e 4.14.

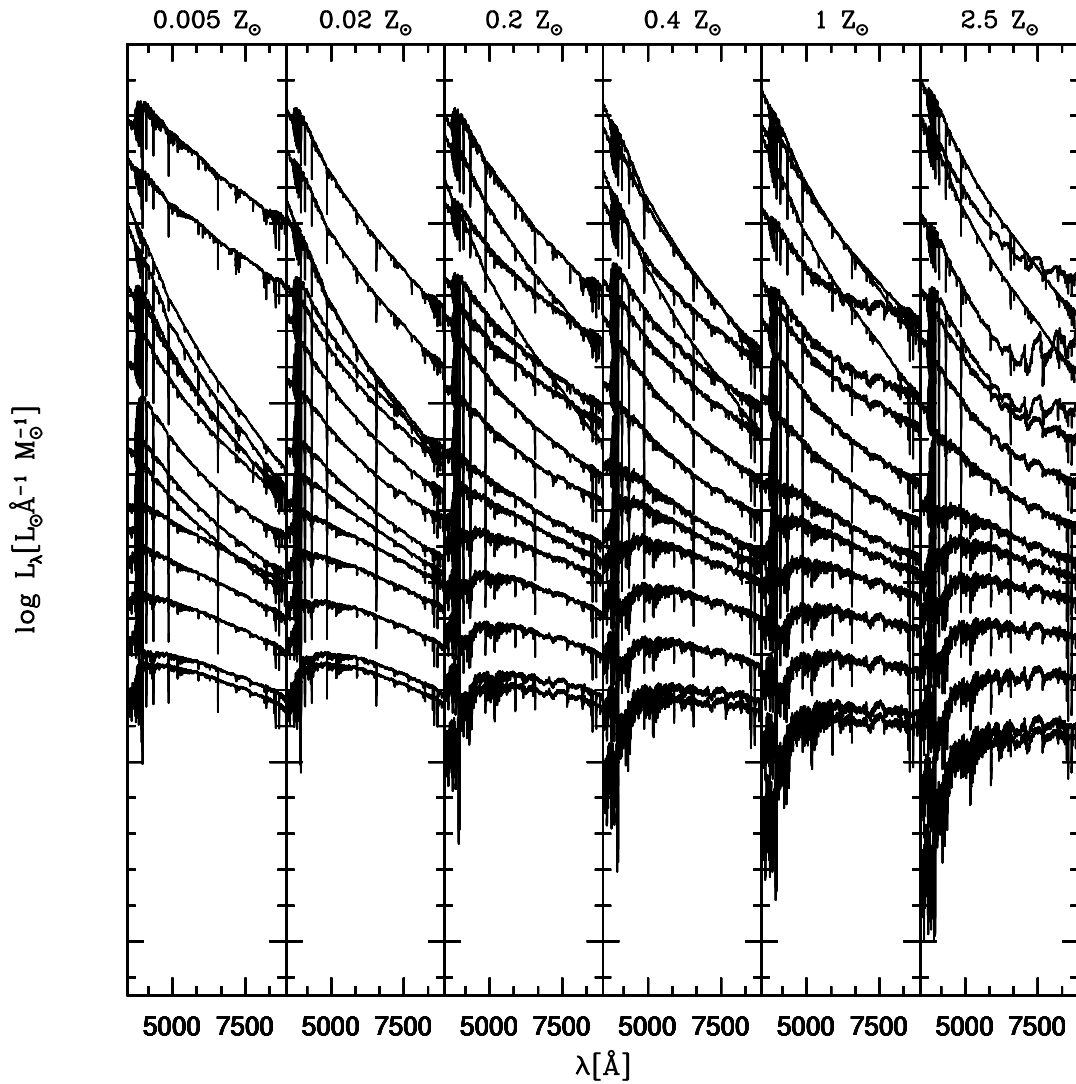


Fig. 4.3: **Base O**: Espectros de 90 populações estelares simples usados na síntese espectral (de BC03). A base compreende 6 diferentes metalicidades:  $Z_{\star} = 0.005, 0.02, 0.2, 0.4, 1$  e  $2.5 Z_{\odot}$ , e 15 idades. De cima para baixo,  $t_{\star} = 0.001, 0.00316, 0.00501, 0.01, 0.02512, 0.04, 0.10152, 0.28612, 0.64054, 0.90479, 1.434, 2.5, 5, 11$  e  $13 \times 10^9$  anos.

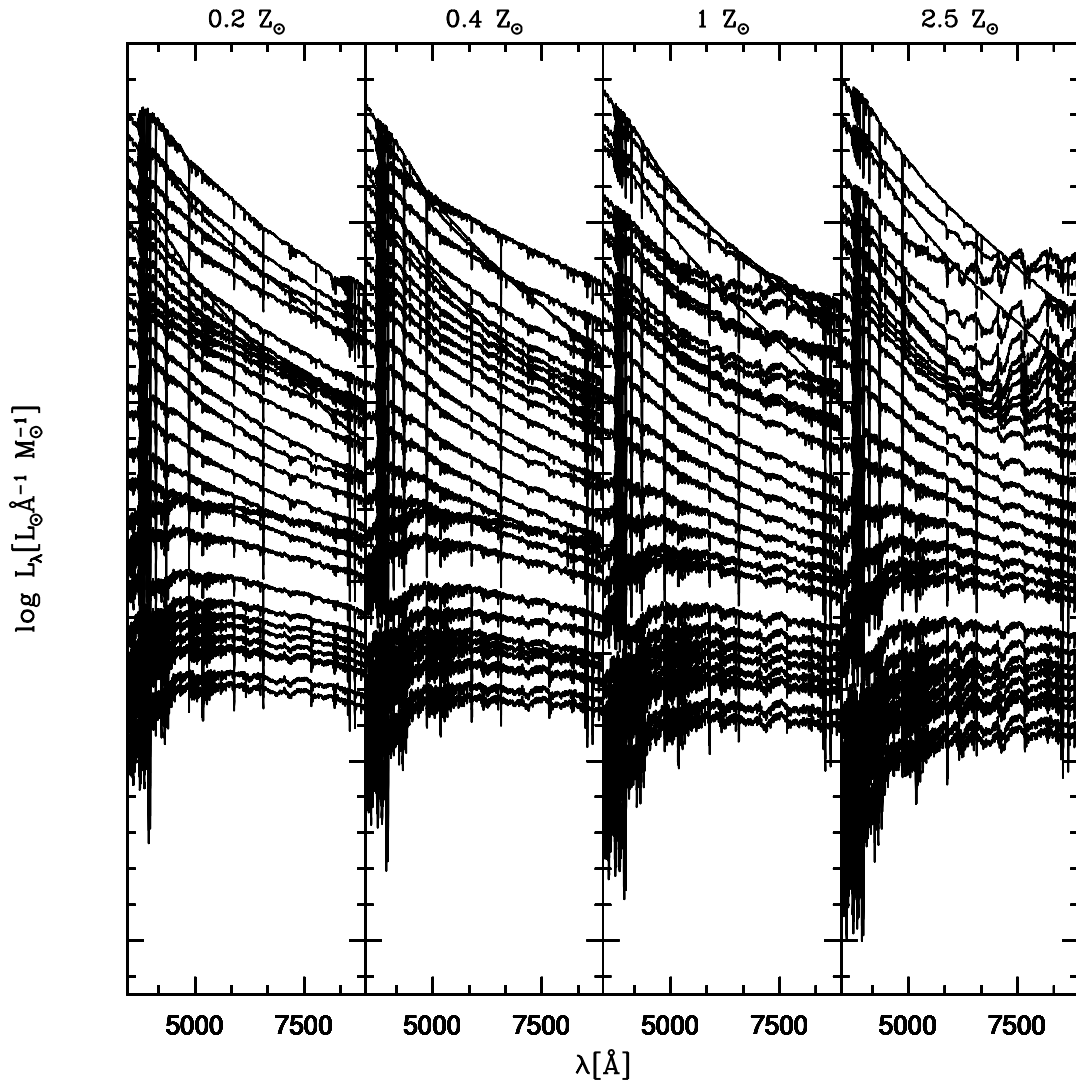


Fig. 4.4: **Base P:** Espectros de 120 populações estelares simples usados na síntese espectral (de BC03). A base compreende 4 diferentes metalicidades:  $Z_{\star} = 0.2, 0.4, 1$  e  $2.5 Z_{\odot}$ , e 30 idades. De cima para baixo,  $t_{\star} = 0.001, 0.00316, 0.00501, 0.00661, 0.00871, 0.01, 0.01318, 0.01445, 0.01995, 0.02512, 0.03200, 0.04, 0.055, 0.10152, 0.16090, 0.28612, 0.40415, 0.64054, 0.90479, 1.27805, 1.43400, 2.5, 3.5, 4.25, 5, 6.25, 7.5, 7.75, 11$  e  $13 \times 10^9$  anos.

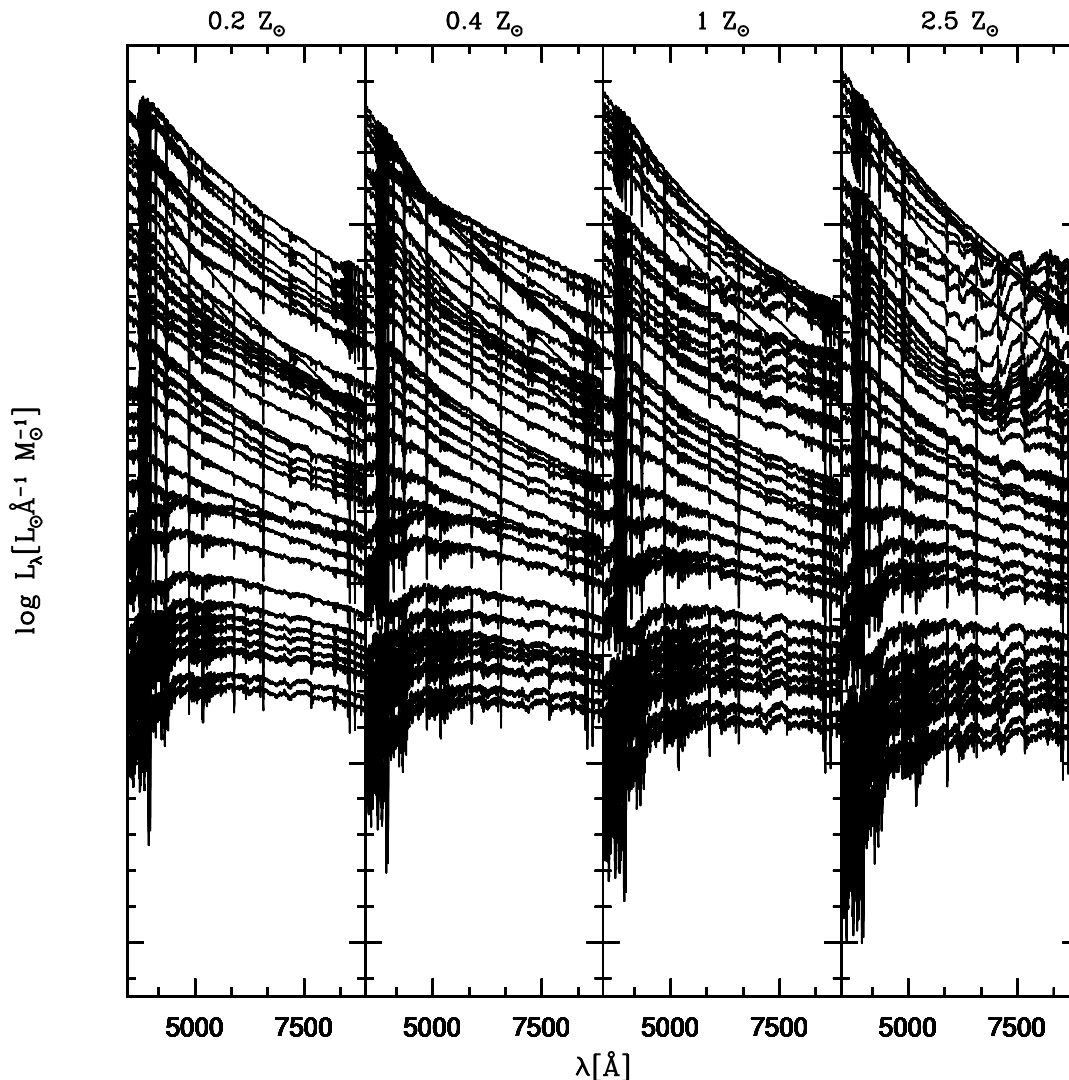


Fig. 4.5: **Base Q:** Espectros de 152 populações estelares simples usados na síntese espectral (de BC03). A base compreende 4 diferentes metalicidades:  $Z_{\star} = 0.2, 0.4, 1.0$  e  $2.5 Z_{\odot}$ , e 38 idades. De cima para baixo,  $t_{\star} = 0.001, 0.00251, 0.00316, 0.00380, 0.00417, 0.00501, 0.00550, 0.00661, 0.00724, 0.00871, 0.01, 0.01318, 0.01445, 0.01738, 0.01995, 0.02512, 0.032, 0.04, 0.055, 0.10152, 0.11391, 0.16090, 0.20256, 0.28612, 0.40415, 0.64054, 0.90479, 1.27805, 1.434, 2.5, 3.5, 4.25, 5, 6.25, 7.5, 7.75, 11$  e  $13 \times 10^9$  anos.

metallicidade e forçamos que o modelo tenha uma metalicidade maior (devido à base utilizada), fazemos com que a idade da galáxia diminua para se ajustar ao espectro, já que idades mais jovens possuem espectros mais azuis, resultando assim em uma concentração de galáxias em um lugar que não deveriam estar (por isso o duplo pico).

Vale a pena enfatizar que esse não é um problema restrito ao método. Ele também aparece quando utiliza-se  $D_n4000$  (figura 4.7), que é um “relógio” empírico muito usado atualmente (K03a, b e c). Sabemos, pelo capítulo 3, que a idade cresce à medida que  $D_n4000$  cresce. Podemos notar que temos um duplo pico, sendo o primeiro, que corresponde a idade média de  $10^{9.2}$  anos, para  $D_n4000 \sim 1.5$ .

Se graficarmos o  $D_n4000$  versus a idade para populações estelares simples, fixando-se a metalicidade, vemos que para o valor do pico em  $D_n4000 \sim 1.5$  desloca-se para idades mais altas quando introduzimos metalicidades mais baixas. Esse gráfico pode ser visto na figura 4.8. Isto implica que precisamos de um conjunto de metalicidades mais baixas para acabarmos com o problema do duplo pico.

Conclui-se que temos um problema a menos, resolvido com a base O que introduz metalicidades mais baixas na base.

### 4.3 O “Triângulo das Bermudas”

Se por um lado a inclusão de baixas metalicidades conserta o problema da distribuição de  $\langle \log t_\star \rangle_M$  (e de  $\langle \log t_\star \rangle_L$ ), o mesmo não acontece com o segundo “problema” mencionado que, com a mudança de bases, ainda persiste. Podemos ver na figura 4.9 que na região de idades altas e metalicidades altas ainda temos falta de galáxias no triângulo superior direito desse diagrama. Portanto, o efeito “*triângulo das bermudas*” continua existindo.

A distribuição das galáxias nessa região é na direção da degenerescência entre idade e metalicidade, que deve ter um papel mais severo nessa região pois as idades das galáxias são maiores. Porém, a amplitude do problema parece ser bem maior do que nas simulações realizadas no capítulo 2 (figura 2.10).

Poderíamos fazer uma vistoria dos ajustes dos espectros para avaliarmos a qualidade dos mesmos. Ao contrário do que se poderia imaginar, galáxias ao longo da linha que vai de  $(\langle \log t_\star \rangle_L, \log \langle Z_\star \rangle_L) = (13 \times 10^9 \text{ anos}, 1.0 Z_\odot)$  até  $(1.434 \times 10^9 \text{ anos}, 2.5 Z_\odot)$  são bem ajustadas pelo método, como mostram as figuras 4.12 – 4.14. As melhoras no desvio percentual médio entre modelo e espectro observado  $\Delta_\lambda$  em relação às novas bases são modestos, da ordem de  $\sim 0.1\%$  (figura 4.10). Em termos do ajuste, fica difícil decidir qual base é a mais adequada. Dentro da faixa de idades e metalicidades varridas pelas bases  $\mathcal{B}_N$ ,  $\mathcal{B}_O$ ,  $\mathcal{B}_P$  e  $\mathcal{B}_Q$ , não há uma diferença significativa em relação ao uso das mesmas.

Esse, então, é um problema que não conseguimos resolver e tão pouco entender qual a sua causa.

### 4.4 Simulações: O Triângulo das Bermudas

A primeira pergunta que podemos fazer é: “Será que há algum problema com

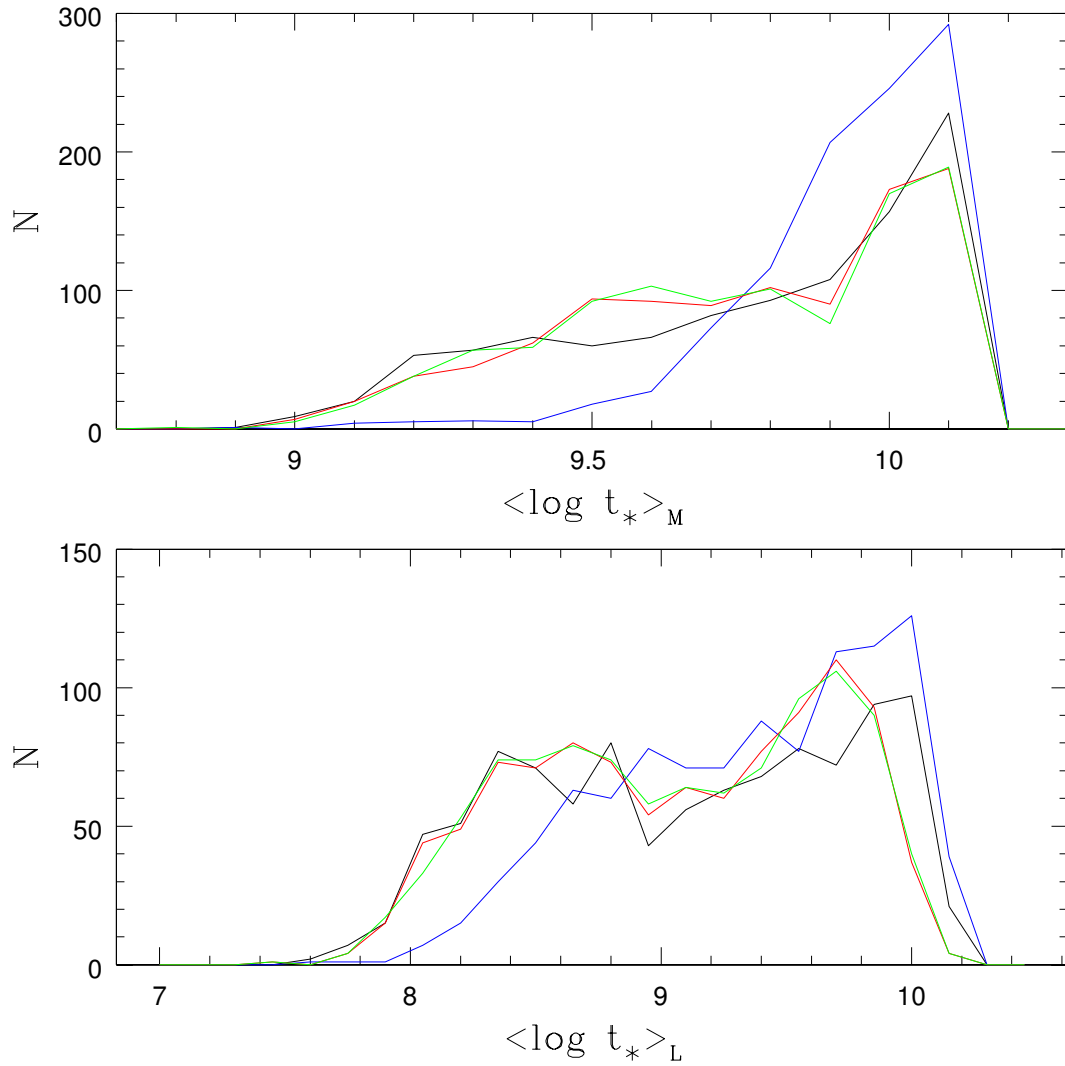


Fig. 4.6: Histogramas da idade média ponderada pela massa e pelo fluxo para a amostra randômica de 1000 galáxias da SDSS sintetizadas com a  $\mathcal{B}_N$  (preta),  $\mathcal{B}_O$  (azul),  $\mathcal{B}_P$  (vermelha) &  $\mathcal{B}_Q$  (verde). Podemos ver que o duplo pico deixa de existir com a base O. Isto é devido a metalicidades mais baixas serem mais “azuis” (mais quentes) para um mesmo conjunto de idades do que metalicidades mais altas.

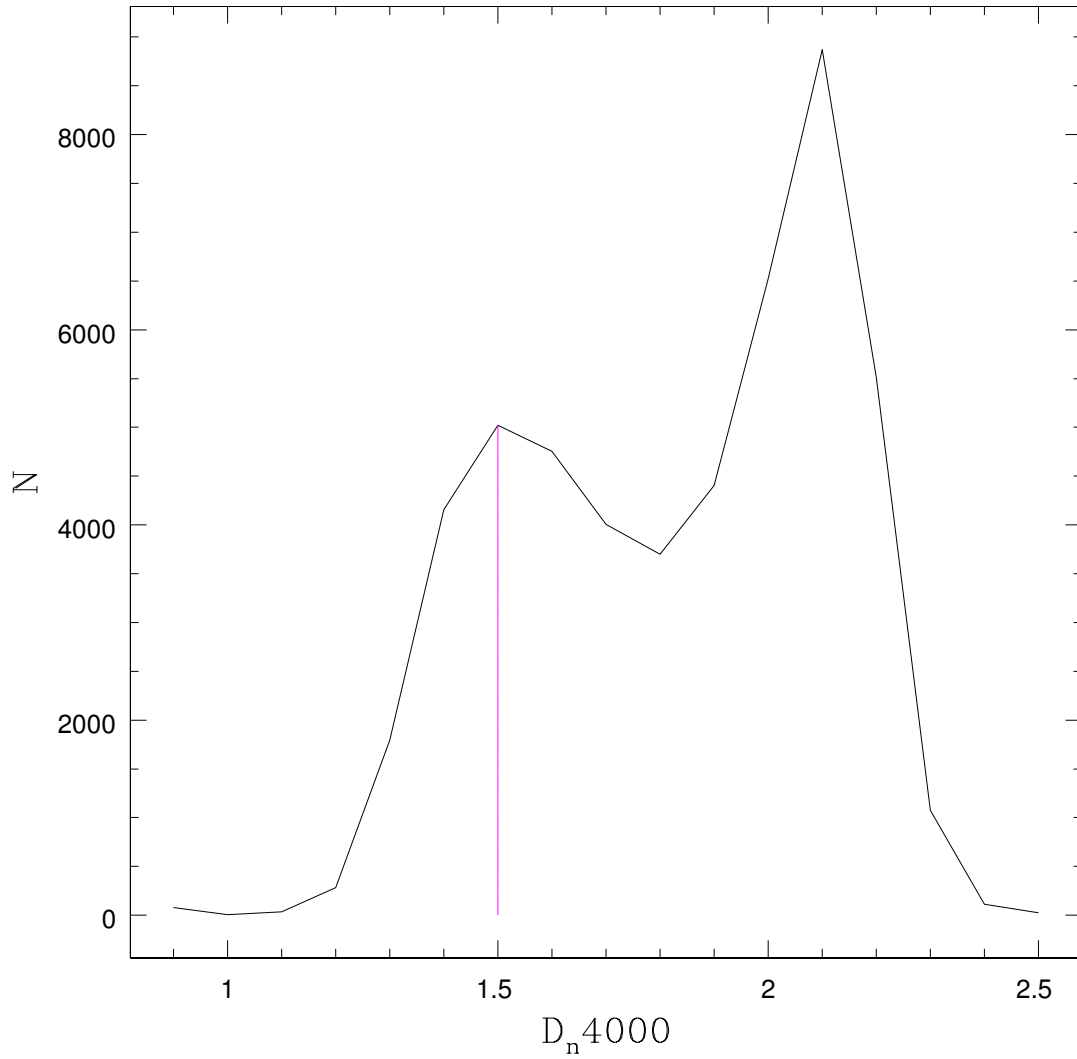


Fig. 4.7: Histograma do  $D_n4000$  para a amostra de 50362 galáxias da SDSS sintetizadas com a  $\mathcal{B}_N$ . Podemos ver uma curva com um duplo pico. O primeiro, com  $D_n4000 = 1.5$ , corresponde a idade estelar média ponderada pela massa de  $\sim 10^{9.2}$  anos (ver figura 4.1).



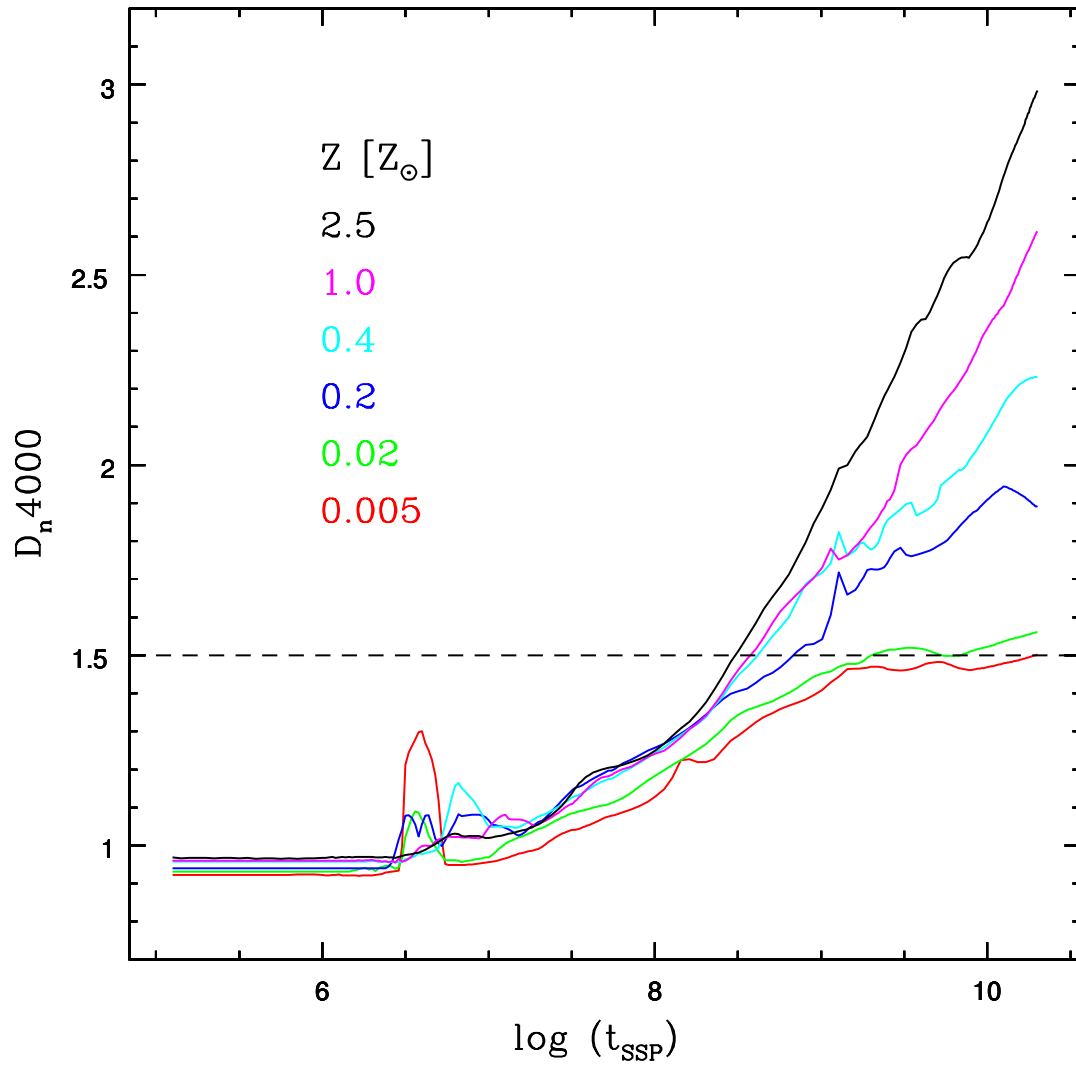


Fig. 4.8:  $D_n 4000$  versus idades de populações estelares simples para cada uma das metalicidades dadas:  $0.005 Z_\odot$  (vermelho),  $0.02 Z_\odot$  (verde),  $0.2 Z_\odot$  (azul),  $0.4 Z_\odot$  (ciano),  $1.0 Z_\odot$  (magenta) e  $2.5 Z_\odot$  (preto). A linha tracejada indica o pico em  $D_n 4000 = 1.5$  observado em galáxias da SDSS (figura 4.7).

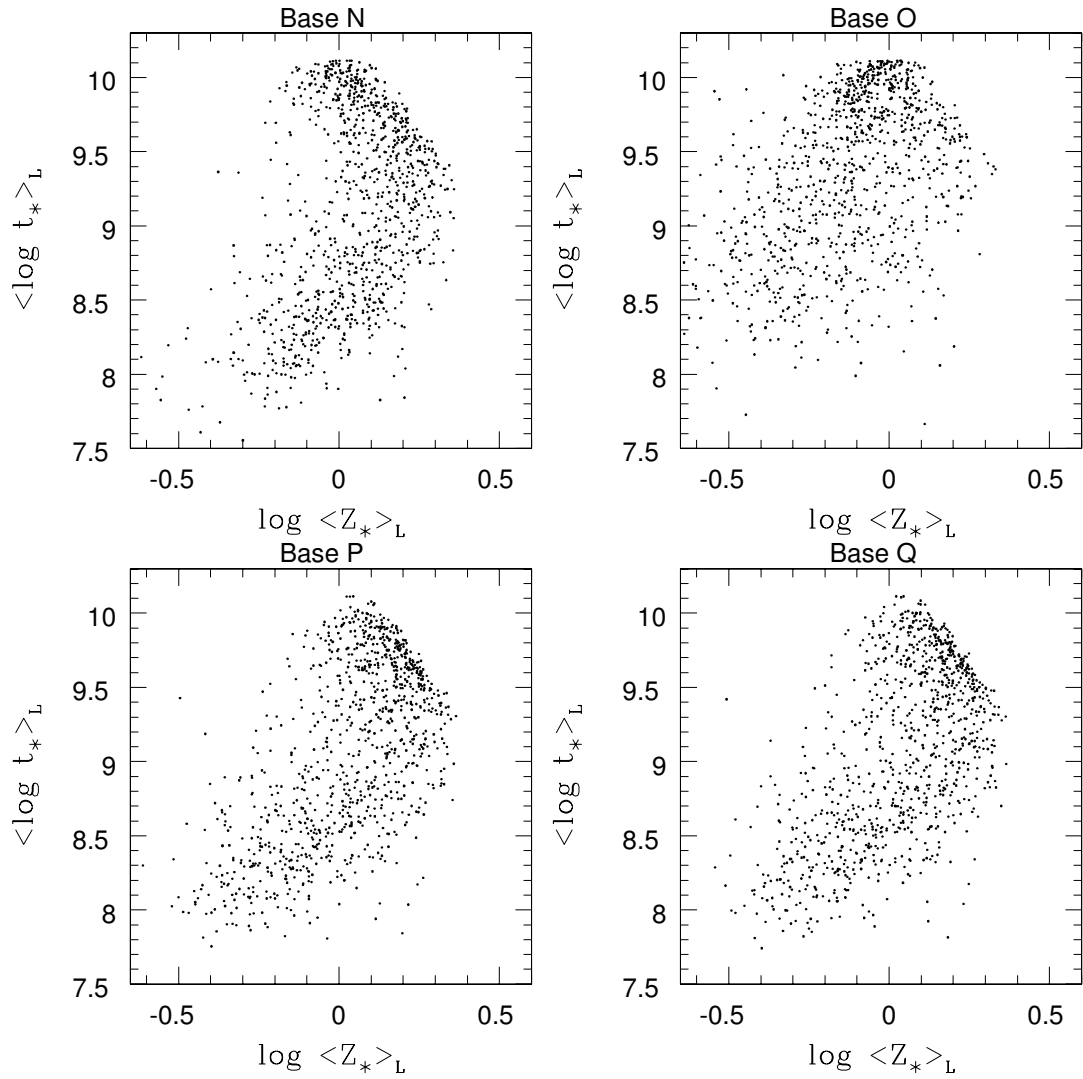


Fig. 4.9: Amostra aleatória de 1000 galáxias comparando idade média contra metalicidade média pesada pelo fluxo obtidas com as bases  $\mathcal{B}_N$ ,  $\mathcal{B}_O$ ,  $\mathcal{B}_P$  &  $\mathcal{B}_Q$ .

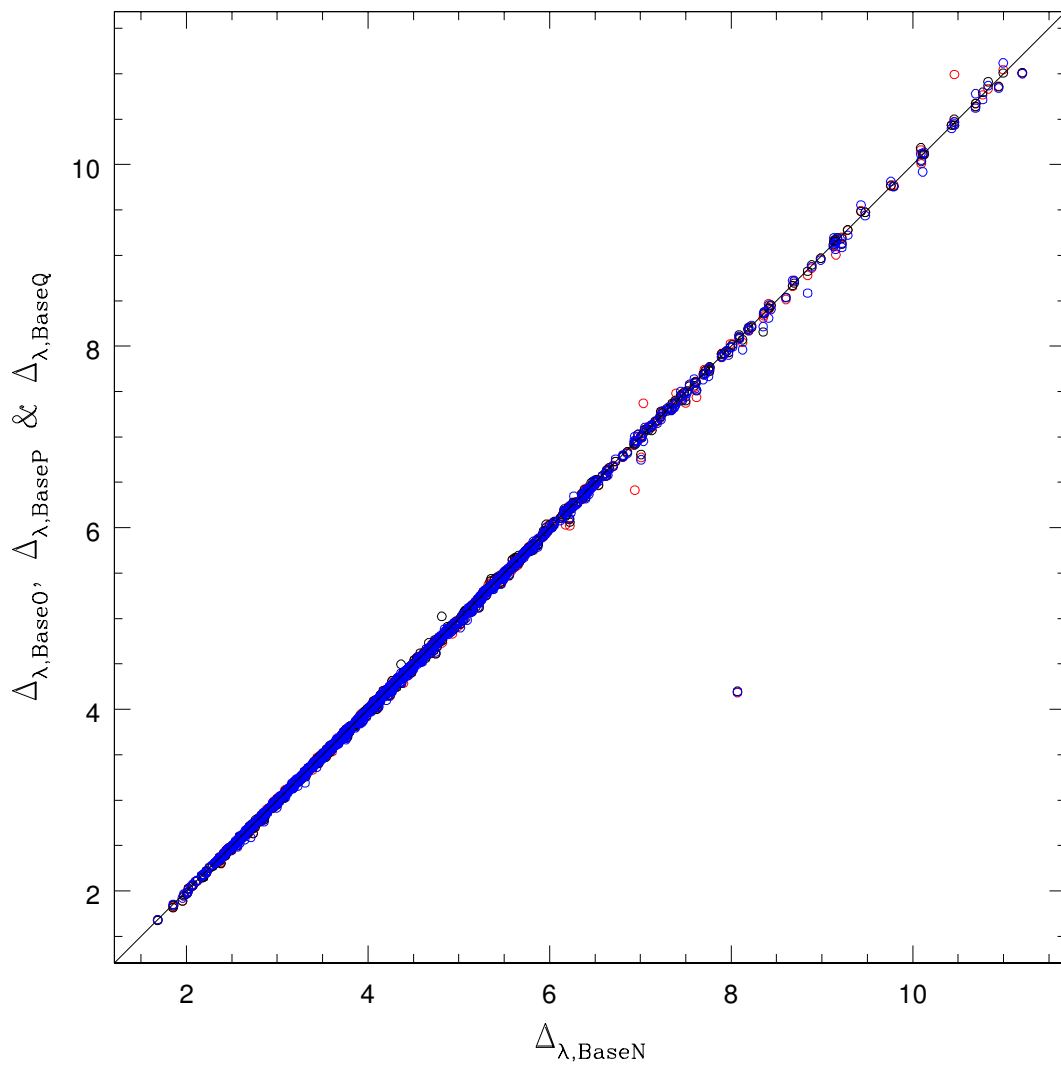


Fig. 4.10: Amostra aleatória de 1000 galáxias comparando os desvios percentuais em relação ao modelo ( $\Delta_{\lambda}$ ) das bases O, P, Q com a base N.

o **STARLIGHT?**” Ou seja, será que quando temos uma galáxia com alta idade e metalicidade (dentro do “triângulo das bermudas”) ela é deslocada para outro ponto do diagrama  $\langle \log t_{\star} \rangle_L$  versus  $\log \langle Z_{\star} \rangle_L$ ? Essa é a motivação para a nova série de simulações descrita abaixo, que tenta responder esta pergunta.

Para investigarmos o efeito “triângulo das bermudas” geramos 85 galáxias teste no quadrante superior direito do diagrama  $\langle \log t_{\star} \rangle_L$  versus  $\log \langle Z_{\star} \rangle_L$  com  $A_V = 0$ . Os limites desse quadrante são:  $1 \leq \langle Z_{\star} \rangle_L \leq 2.5$  e  $8.9 \leq \langle \log t_{\star} \rangle_L \leq 10.2$ . Foram realizadas 10 perturbações utilizando a curva de erro característica da SDSS (figura 2.7) para cada galáxia individual, com  $S/N = 10, 20$  e  $30$  em torno de  $\lambda_0 = 4020\text{\AA}$ . A síntese foi feita com o código **STARLIGHT03** (descrito em 2.1) a fim de tentar recuperar os parâmetros de entrada (vetor de populações, extinção e dispersão de velocidade).

Conseguimos recuperar os valores de entrada com erros de  $-0.2 < \Delta \langle \log t_{\star} \rangle_L < 0.2$  dex para a idade estelar média para  $S/N > 10$ ; para a metalicidade estelar média erros de  $-0.16 < \Delta \log \langle Z_{\star} \rangle_L < 0.16$  para  $S/N > 10$  (figura 4.11). Mostramos assim um resultado semelhante ao obtido com as simulações no capítulo 2, porém aqui para o “triângulo das bermudas”. Estes resultados indicam que não há problemas na obtenção dos parâmetros para uma galáxia que se encontra na região de altas metalicidades e altas idades.

Conclui-se que o código não tem nenhum problema em recuperar os parâmetros de entrada nessa região (dentro dos limites impostos pela degenerescência entre idade e metalicidade). Isto já era esperado pois estas simulações são similares às realizadas no capítulo 2 com a diferença de que aqui nos concentramos na análise de galáxias com idades e metalicidades estelares médias altas.

## 4.5 Quem São Essas Galáxias?

Com as discrepâncias ainda não esclarecidas, investigamos se algum grupo particular de galáxias encontra-se nessa estranha linha (figura 4.2). Por simples inspeção dos resultados anteriores, esperamos que estas sejam as galáxias elípticas (idades e metalicidades estelares médias elevadas). Para verificarmos nossa suposição, fizemos a análise dessas por meio de alguns critérios como: dispersões de velocidades (possuem  $\sigma_{\star}$  elevado), índice de concentração (elípticas possuem índice de concentração maior que 2.6) e cores (galáxias com espectros vermelhos e com poucas linhas de emissão). Estas são características de galáxias elípticas pois possuem populações estelares velhas e muito pouco gás.

### 4.5.1 Resíduos Espectrais

Ao analisarmos de perto os espectros dessas galáxias que seguem a linha na borda superior direita do gráfico 4.2, vemos que, individualmente (figuras 4.12, 4.13 e 4.14) os ajustes são excelentes. Porém, é possível notar resíduos ( $O_{\lambda} - M_{\lambda}$ ) sistemáticos em algumas bandas.

Para tornar mais aparente os resíduos, procuramos realizar uma média sobre os espectros  $O_{\lambda}$  de elípticas, onde para cada comprimento de onda  $\lambda$  temos um número

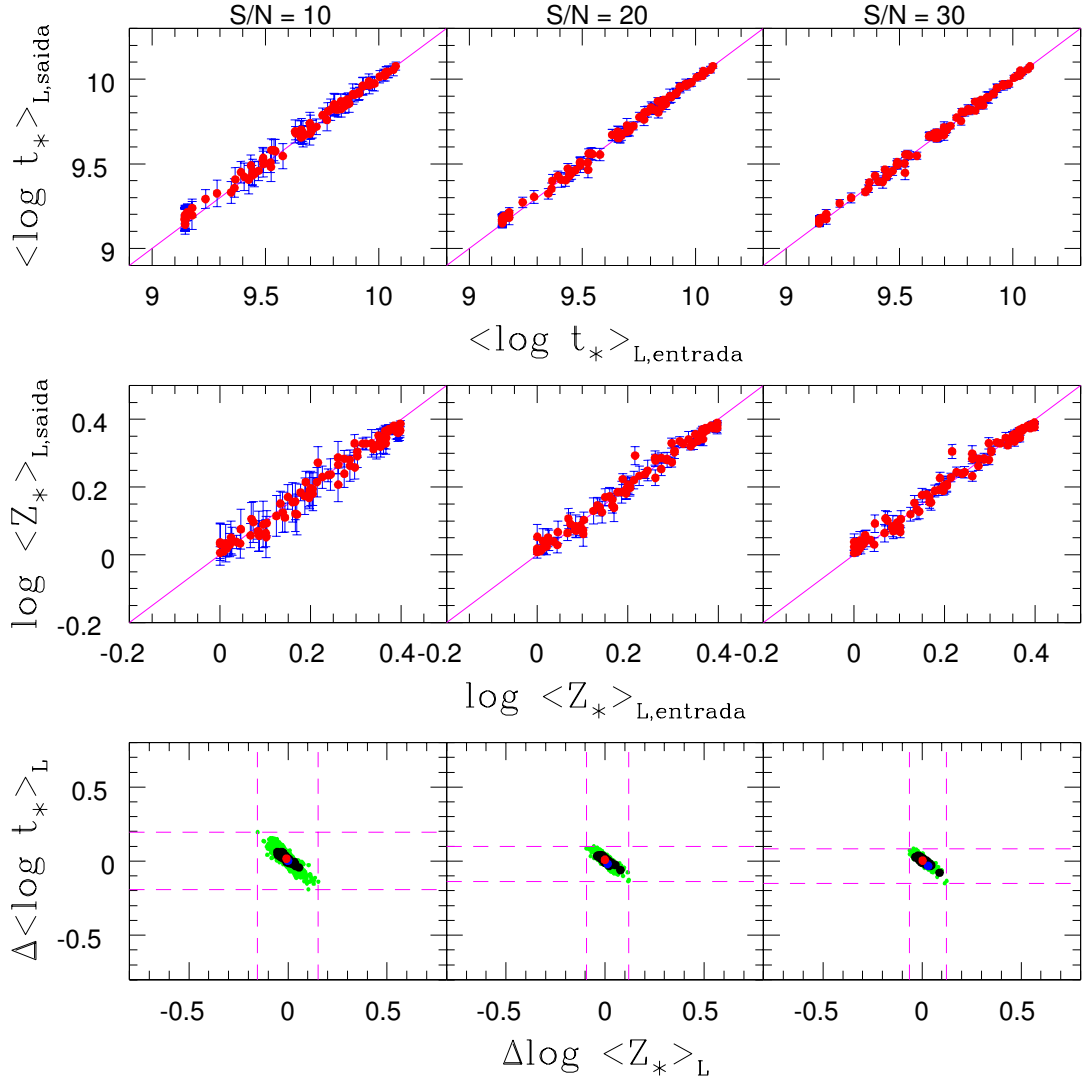


Fig. 4.11: Idade estelar média e metalicidade estelar média de entrada versus saída e degenerescência entre idade e metalicidade para 85 galáxias teste geradas e que caem no quadrante superior direito do diagrama idade estelar média contra metalicidade estelar média ( $0 \leq \log \langle Z_* \rangle_L \leq 0.4$  e  $8.9 \leq \langle \log t_* \rangle_L \leq 10.2$ ). Cada galáxia teste foi perturbada 10 vezes com  $S/N = 10, 20$  e  $30$  em torno de  $\lambda_0 = 4020 \text{ \AA}$  (utilizou-se o mesmo espectro de erro característico da SDSS da figura 2.7).

*Idade estelar média (painel superior) ou metalicidade estelar média (painel do meio) de entrada versus saída:* Pontos são as médias das 10 versões perturbadas com as suas respectivas barras de erro. A linha cheia mostra onde deveriam cair os pontos idealmente, caso a síntese recuperasse com perfeição os valores de entrada.

*Painel inferior: Diagrama de degenerescência entre idade e metalicidade média:* Pontos com cores verdes são as 850 versões perturbadas das galáxias teste e pontos em preto são as médias das 10 versões perturbadas de cada galáxia.

de pontos  $N_{\lambda,OK}$  que corresponde ao número de galáxias que nesse comprimento de onda não possuem píxeis espúrios:

$$\langle O_{\lambda} \rangle = \frac{\sum_{g=1}^{N_{\lambda,OK}} O_{g,\lambda}}{\sum_{g=1}^{N_{\lambda,OK}} g} \quad (4.1)$$

A equação acima é nada mais do que a média aritmética no comprimento de onda  $\lambda$ . Obtivemos também, a média dos resíduos:

$$\langle O_{\lambda} - M_{\lambda} \rangle = \frac{\sum_{g=1}^{N_{\lambda,OK}} (O_{g,\lambda} - M_{g,\lambda})}{\sum_{g=1}^{N_{\lambda,OK}} g} \quad (4.2)$$

e a raiz quadrada do resíduo quadrático médio:

$$\langle (O_{\lambda} - M_{\lambda})^2 \rangle^{1/2} = \left( \frac{\sum_{g=1}^{N_{\lambda,OK}} (O_{g,\lambda} - M_{g,\lambda})^2}{\sum_{g=1}^{N_{\lambda,OK}} g} \right)^{1/2} \quad (4.3)$$

Podemos ver que, ao realizarmos essas médias para  $\sim 2000$  elípticas, os resíduos tornam-se mais aparentes (figura 4.15). Os resíduos mais gritantes se encontram em  $\lambda \sim 4200, 5170$  e  $5890 \text{ \AA}$ , que correspondem a linhas de absorção associadas ao CN, Mg e Na, respectivamente. O modelo  $M_{\lambda}$  está sendo sistematicamente subestimado nestas bandas.

Encontramos dois resíduos sistemáticos em  $\sim 6900 \text{ \AA}$  e  $\sim 7600 \text{ \AA}$ . Não nos demos conta de que estes resíduos eram relatados em BC03<sup>2</sup> e, portanto, estes não foram previamente mascarados para a síntese. Para todas as simulações posteriores é conveniente que continuemos a não mascará-los para obtermos uma comparação mais justa.

#### 4.5.2 Extinção Negativa

Outra peculiaridade é que encontramos uma extinção levemente menor do que zero para todos os casos. Isto indica que precisamos de componentes mais azuis, isto é, de estrelas mais quentes, do que as disponíveis em nossa base.

O código permitia, em todas as simulações anteriores, que a extinção fosse negativa. A justificativa para isso é puramente estatística, pois podemos mapear a função distribuição de  $A_V$  sem truncá-la. Porém, fisicamente, uma galáxia sempre terá uma extinção  $A_V \geq 0$ . Em princípio, isto não parece ser um motivo de preocupação. Contudo, o fato de que galáxias ao longo da linha divisória no diagrama  $\langle \log t_{\star} \rangle_L$  versus  $\log \langle Z_{\star} \rangle_L$  exibam  $A_V < 0$ , e não uma distribuição centrada em 0, indica que algo de estranho está ocorrendo.

## 4.6 Galáxias Elípticas

Como concluímos que as galáxias ao longo dessa linha “problemática” são elípticas, nessa seção fazemos uma pequena pausa para revisar algumas propriedades

<sup>2</sup> Duas regiões com problemas devido a efeitos telúricos são relatadas em BC03: 6850–6950  $\text{\AA}$  e 7550–7725  $\text{\AA}$ .

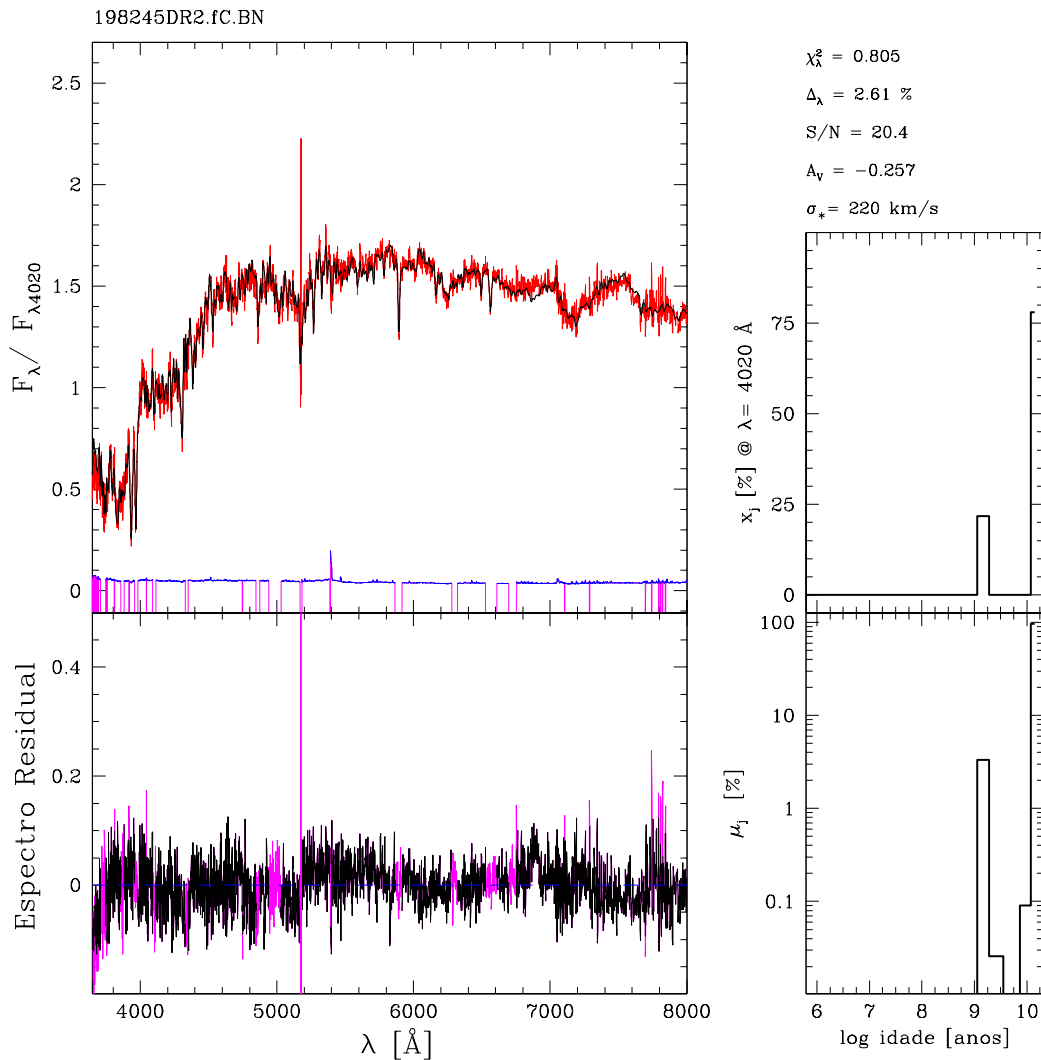


Fig. 4.12: Galáxia que está marcada com a letra A no diagrama 4.2. Podemos ver que o ajuste é muito bom. O objetivo aqui e nas próximas duas galáxias (B e C) é mostrar como individualmente os ajustes são excelentes e que é muito difícil notar algum problema. Agora, fazendo uma média sobre galáxias que encontram-se nessa linha é possível suavizar os resíduos e notar tendências de subestimação em determinadas bandas (como mostramos na figura 4.15).

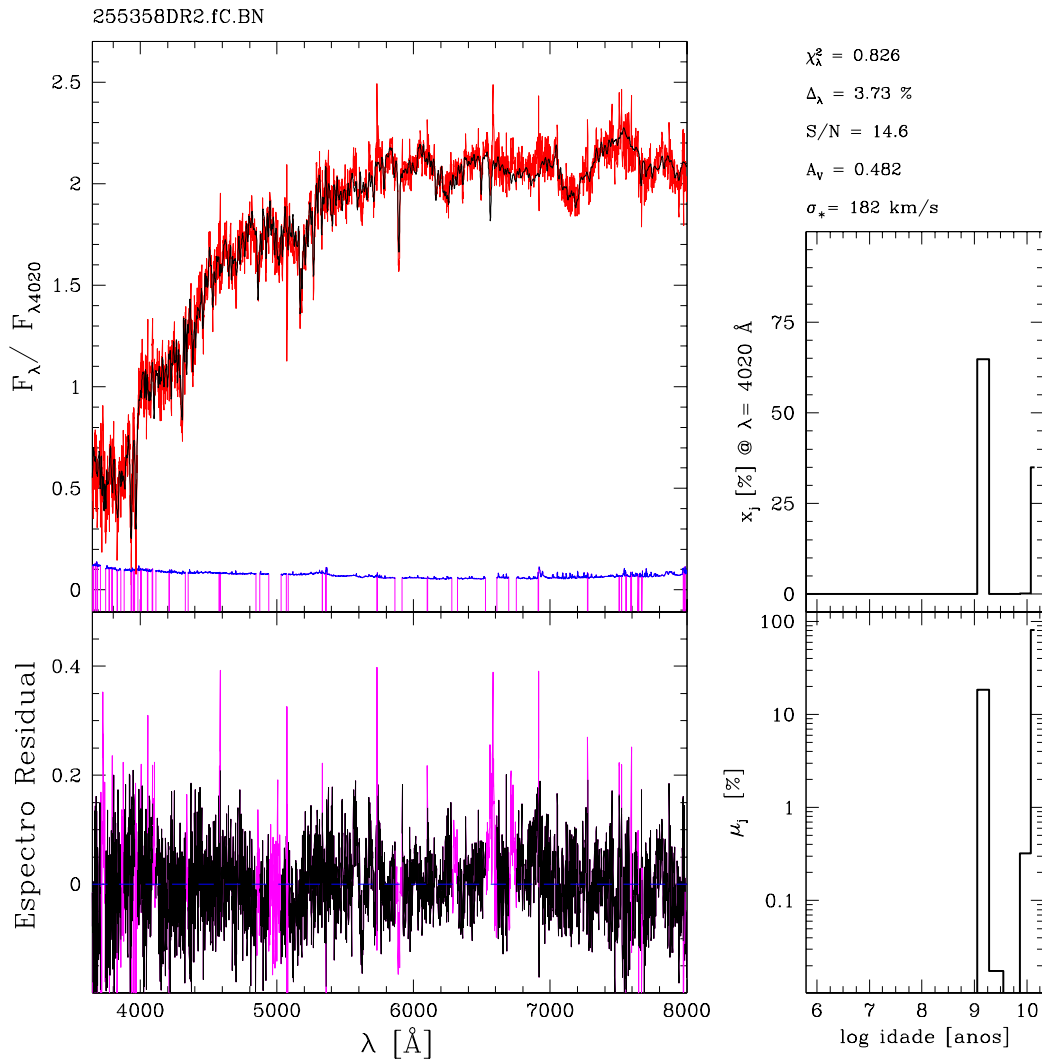


Fig. 4.13: Galáxia que está marcada com a letra B no diagrama 4.2. Podemos ver que o ajuste também é muito bom.



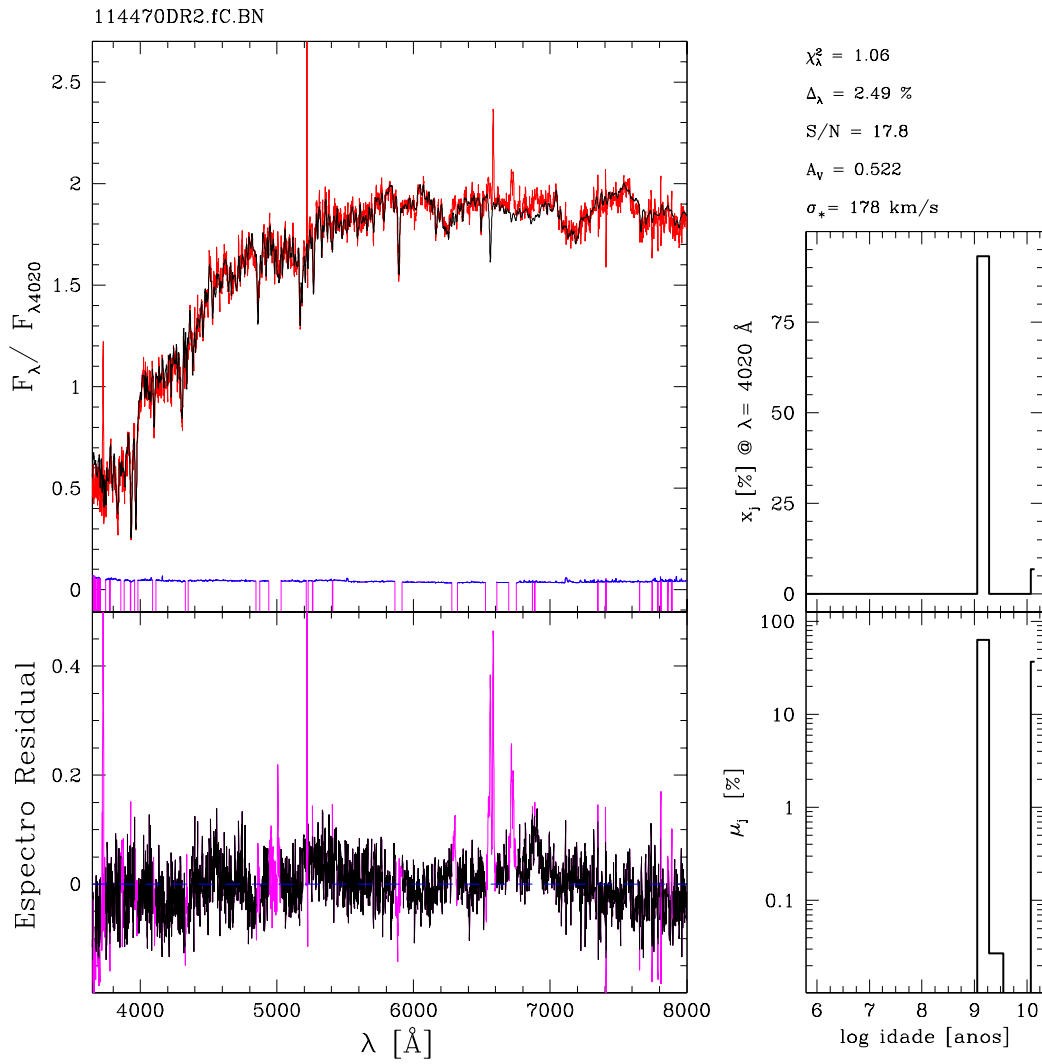


Fig. 4.14: Galáxia que está marcada com a letra C no diagrama 4.2. Nota-se que estas galáxias possuem cores bem avermelhadas sem muitas linhas de emissão e dispersões de velocidades altas, propriedades típicas de galáxias elípticas.

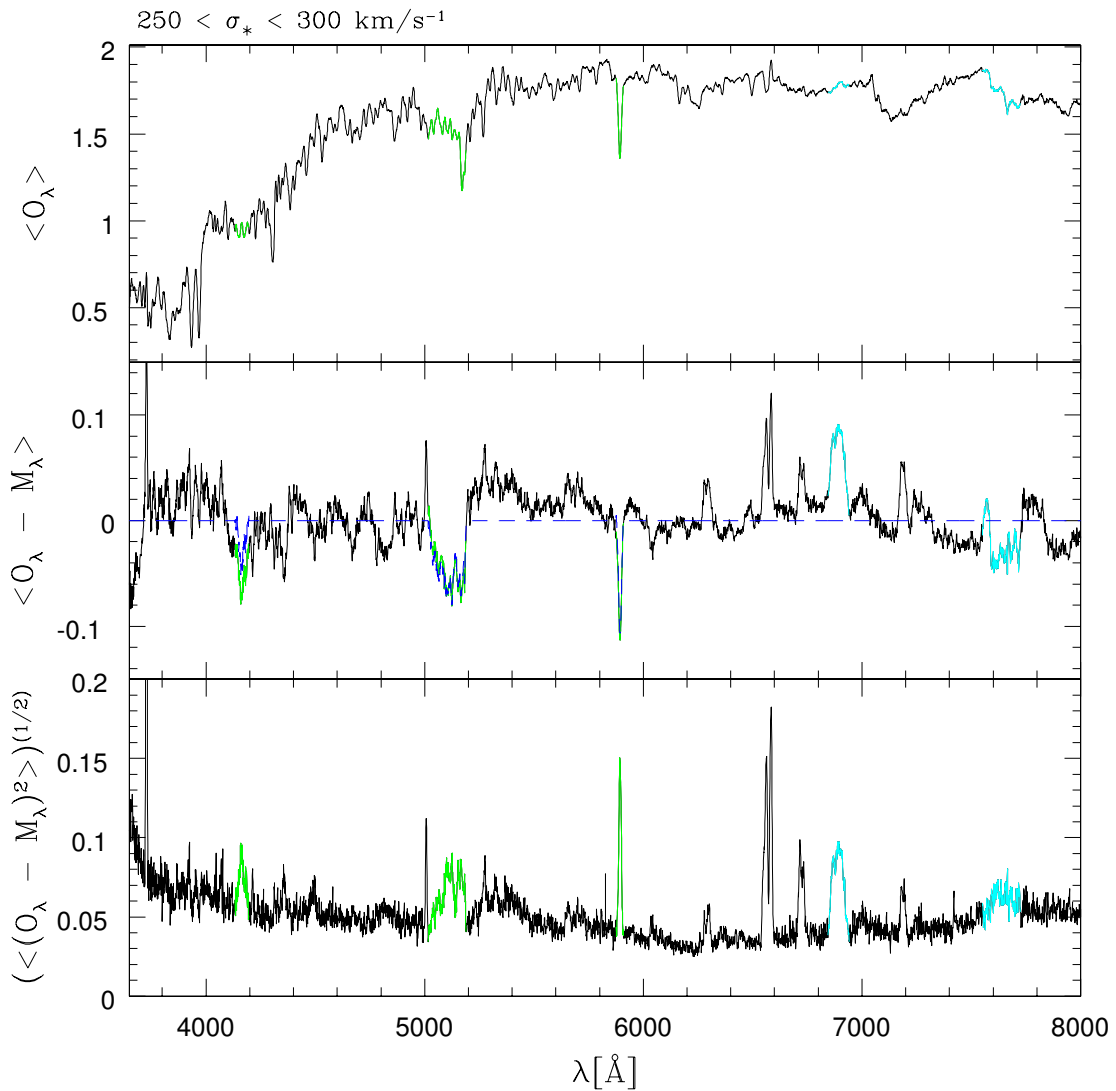


Fig. 4.15: Média dos espectros, resíduos e resíduos quadráticos de  $\sim 2000$  galáxias situadas na estranha linha do quadrante superior direito do diagrama  $\langle \log t_{\star} \rangle_L$  contra  $\log \langle Z_{\star} \rangle_L$  (figura 4.2). As médias foram calculadas selecionando galáxias com dispersão de velocidade entre  $250 \leq \sigma_{\star} \leq 300 \text{ km/s}^{-1}$ . As regiões marcadas em verde da esquerda para a direita nos gráficos são as bandas CN, Mg e Na, respectivamente (definidas nesse trabalho em 4132–4196 Å, 5015–5190 Å e 5874–5911 Å). As regiões marcadas em ciano são um problema relatado em BC03, devido a efeitos telúricos. A linha pontilhada em azul é um espectro residual médio levando-se em conta somente as subestimações sistemáticas nas bandas CN, Mg e Na (esse resíduo será usado para um conjunto de simulações a fim de testarmos o efeito de enriquecimento químico de elementos  $\alpha$ ).

dessas galáxias.

Os primeiros trabalhos de populações estelares em galáxias (Baade 1944) mostravam um cenário razoavelmente simples para galáxias elípticas, onde as populações dessas podem ser modeladas apenas como velhas (praticamente coevas), semelhantes a aglomerados globulares de idades  $\sim 15 \times 10^9$  anos, metalicidades altas, não possuem discos, livres de gás e poeira. Inclusive na aparência possuem um formato extremamente simples: são elipsóides. No entanto, nosso entendimento sobre estas estava muito aquém do que se acreditava.

Nos últimos 30 anos, o que se entendia sobre galáxias elípticas sofreu mudanças severas. As populações estelares de elípticas de baixa massa parecem ser de intermediárias a velhas, em alguns casos encontram-se discos estelares fracos e em geral as elípticas massivas possuem um meio interestelar quente com temperatura  $T > 10^6$  K (são envoltas por uma coroa de raio X). Outras peculiaridades foram encontradas como um “counterrotating core” que são anomalias nos campos de velocidade nos núcleos dessas galáxias, sinalizando um possível “merger” no passado (a primeira galáxia com essa característica foi NGC 5813 descoberta em 1982 por Efstathiou, Ellis & Carter).

Na análise de linhas de absorção, índices de Lick<sup>3</sup>, como Mg e Fe, não eram recuperados por modelos padrão em comparação com os índices medidos. A primeira interpretação foi dada pelo trabalho de Worthey et al. (1992). Existe uma sobrea-bundância de elementos  $\alpha$  ( $\alpha \equiv$  O, Ne, Mg, Si, S, Ar, Ca, Ti; elementos construídos com partículas núcleos de Hélio mais N e Na.) em relação ao Fe. Recentemente essa interpretação foi corroborada por um estudo de aglomerados globulares (estes possuem uma abundância diferente da solar detectado através de espectroscopia de estrelas individuais) servindo como teste empírico para SSPs (Maraston et. al. 2003). Um outro fato importante é que galáxias com alta dispersão de velocidade e, portanto, altas massas são as mais afetadas por *enriquecimento*  $\alpha/Fe$  (ver figura 4 do trabalho de Thomas, Maraston & Bender 2002 reproduzida na figura 4.16).

Para estudar o efeito de enriquecimento  $\alpha/Fe$  em nossos dados, calculamos os resíduos  $O_\lambda - M_\lambda$  médios em três janelas: 4132–4196 Å, 5015–5190 Å e 5874–5911 Å, correspondentes a bandas do CN, Mg, e Na. Estes resíduos foram divididos pelo  $M_\lambda$  mediano em janelas mais largas (de aproximadamente 100 Å de largura) centradas nestas bandas. Essa divisão nos fornece índices de resíduo na forma de larguras equivalentes que medem o excesso de elementos  $\alpha$  nos dados em relação aos modelos.

Na figura 4.17 mostramos estes resíduos contra a dispersão de velocidade. Nota-

<sup>3</sup> Linhas de absorção são expressas em termos de índices para indicar quão fortes ou fracas elas são. São definidas bandas “características” centradas em algum lugar de interesse (algum elemento) e pseudocontínuos para o vermelho e azul dessa banda. Bandas moleculares (MgH) são expressas em magnitudes, enquanto bandas de elementos atômicos em unidades de Å. Calcula-se então pseudocontínuos (região mais pro azul e vermelho da banda definida em  $\lambda_1$ – $\lambda_2$ ) a partir de  $F_P = \int_{\lambda_1}^{\lambda_2} O_\lambda d\lambda / \int_{\lambda_1}^{\lambda_2} d\lambda$ . Após isto, é definido  $F_{C\lambda}$  que é o fluxo de uma linha reta ligando o ponto médio no contínuo azul ao ponto médio no contínuo vermelho. Sabendo que  $F_{I\lambda}$  é o fluxo por unidade de comprimento de onda para a banda de índice I, definimos largura equivalente ( $[Å]$ ) como  $EW = \int_{\lambda_1}^{\lambda_2} (1 - F_{I\lambda}/F_{C\lambda}) d\lambda$  e um índice medido em magnitudes por  $Mag_I = -2.5 \log[1/(\lambda_2 - \lambda_1) \int_{\lambda_1}^{\lambda_2} F_{I\lambda}/F_{C\lambda} d\lambda]$  (Worthey et al. 1994).

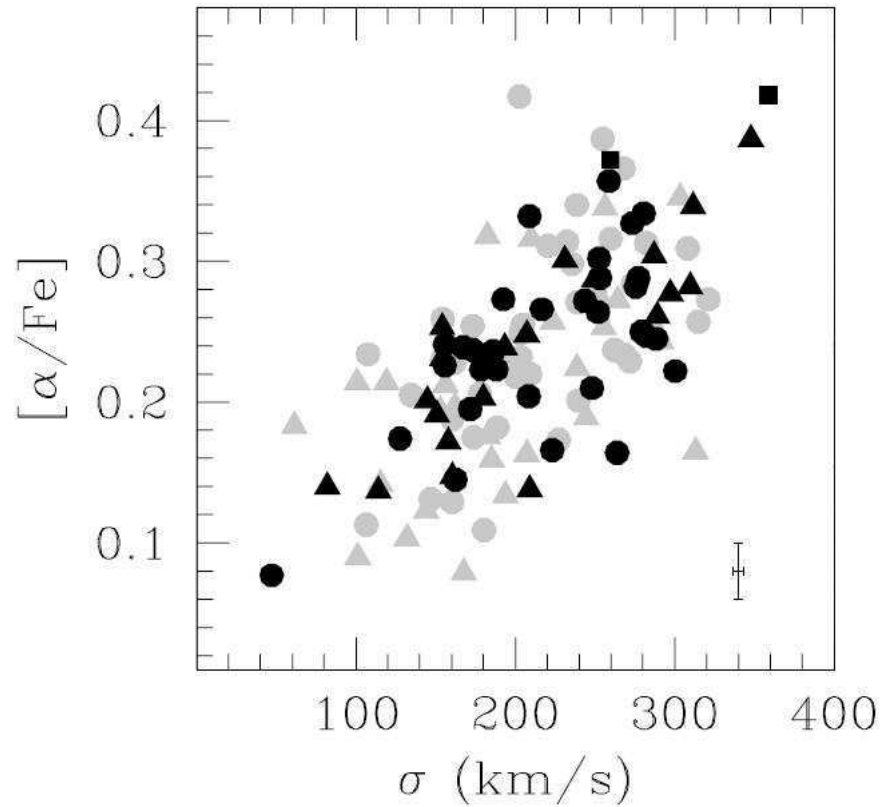


Fig. 4.16: Abundância de elementos  $[\alpha/Fe]$  como função da dispersão de velocidade  $\sigma$ . Símbolos cinzas e pretos denotam galáxias “early-type” do campo e de aglomerados, respectivamente. Triângulos são lenticulares, e círculos são elípticas. Os dois quadrados são as galáxias NGC 4874 e NGC 4889 do aglomerado de Coma. Essa figura é retirada do trabalho de Thomas, Maraston & Bender 2002.

se que o efeito de enriquecimento  $\alpha/Fe$  está fortemente ligado a  $\sigma_*$  confirmando que elípticas massivas são as que apresentam maior  $\alpha/Fe$ .

As diferentes cores ilustram faixas de extinções: azul ( $A_V < -0.2$ ), verde ( $-0.2 \leq A_V \leq 0$ ) e vermelho ( $A_V > 0$ ). Podemos perceber que as galáxias com resíduos maiores têm extinções mais negativas. Uma possível interpretação para esse efeito é que estrelas com enriquecimento  $\alpha/Fe$  são mais quentes do que as escalonadas com abundâncias solares (Salasnich et al. 2000 e Thomas et al. 2003). Assim, quando o código obtém  $A_V < 0$  para uma galáxia, isto é um indicativo de que existe uma mistura química diferente da solar. O código sente a necessidade de “esquentar” (=azular) os espectros da base. Como ele não pode mexer na base, opta por aplicar uma extinção menor do que zero, o que causa um efeito similar. Em outras palavras,  $A_V < 0$  é um efeito colateral de uma base inadequada para modelar galáxias elípticas.

#### 4.7 Simulações: $\alpha/Fe$

Para investigarmos se um enriquecimento  $\alpha/Fe$  causa o efeito “triângulo das bermudas” que tanto nos intriga, realizamos uma nova série de simulações. Adicionamos um resíduo médio em determinadas bandas (CN, Mg, Na) para as mesmas 85 galáxias teste geradas anteriormente (seção 4.4). O objetivo nestas simulações é verificar como o enriquecimento de elementos  $\alpha$  afeta nossos parâmetros. O resíduo médio foi calculado para galáxias com dispersão de velocidade entre  $250 < \sigma_* < 300$  km/s<sup>-1</sup> e que estão situadas nessa estranha linha (figura 4.2). Podemos vê-lo na figura 4.15 no segundo gráfico do resíduo como uma linha tracejada (em azul). Um exemplo de um espectro de uma galáxia teste gerada com o acréscimo desse resíduo pode ser vista na figura 4.18.

As galáxias teste enriquecidas por elementos  $\alpha$  foram sintetizadas com nosso código usando diferentes configurações:

1. Permitindo que a extinção seja negativa.
2. As extinções têm de ser maiores ou iguais a zero.
3. As extinções têm de ser maiores ou iguais a zero e regiões espectrais enriquecidas por elementos  $\alpha$  são previamente mascaradas.

##### 4.7.1 Configuração 1 ( $A_V$ livre)

As simulações com  $A_V$  livre, resumidas na figura 4.19, mostram que temos, para as 850 galáxias teste (85 galáxias vezes 10 perturbações), um resultado sistematicamente de extinções negativas, envelhecimento das galáxias e diminuição da composição química.

Os erros são maiores do que nas simulações anteriores para os valores de entrada (capítulo 2). Além disto, os erros vão no sentido de superestimar os valores da idade média e subestimar os valores da metalicidade média, ou seja, os erros vão desde  $-0.17 < \Delta \langle \log t_* \rangle_L < 0.41$  dex para a idade estelar média e para a metalicidade estelar média de  $-0.17 < \Delta \log \langle Z_* \rangle_L < 0.23$  dex para  $S/N \geq 10$  (figura 4.19).

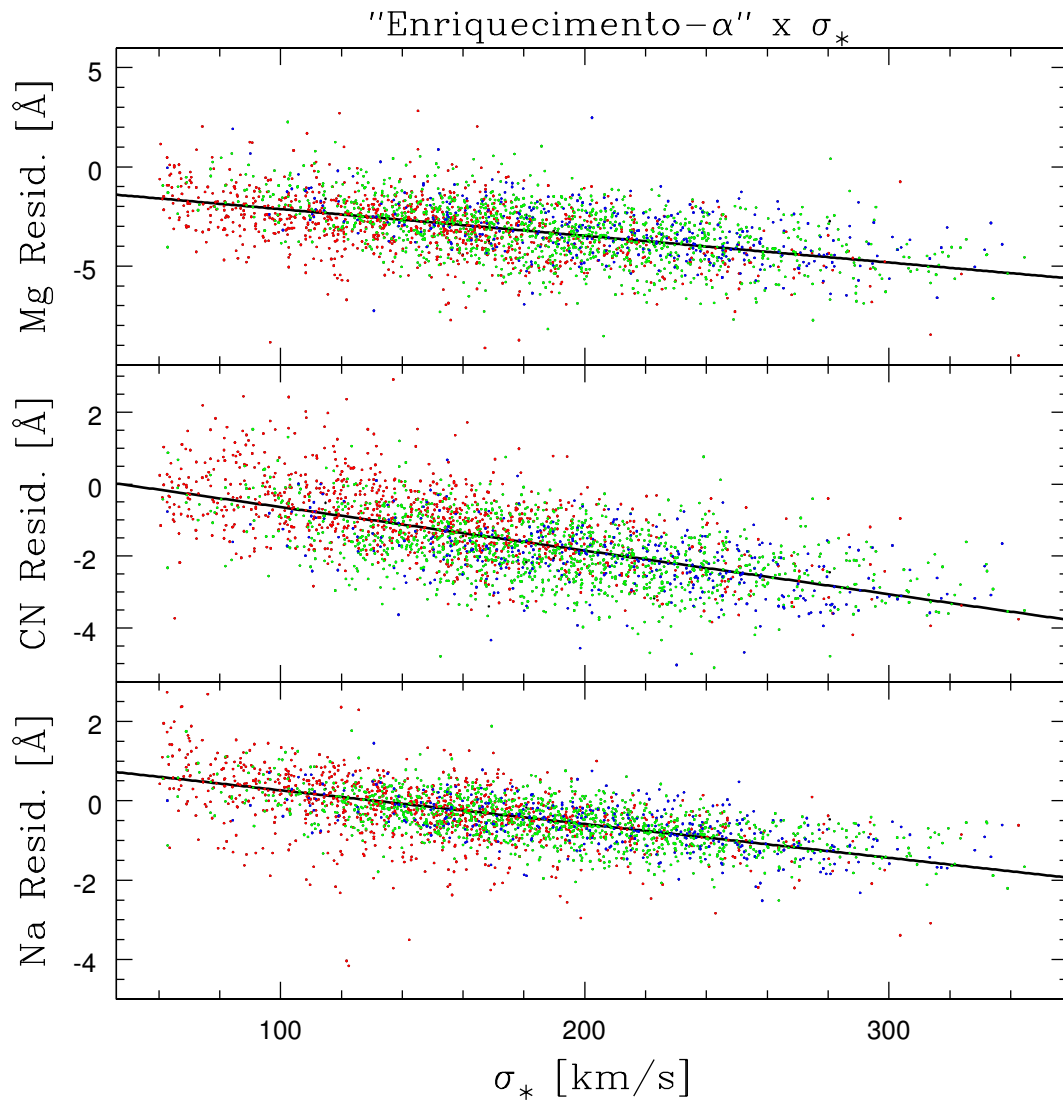


Fig. 4.17: Resíduo das linhas de absorção Mg (5015–5190 Å), CN (4132–4196 Å) e Na (5874–5911 Å) contra  $\sigma_*$ . As galáxias elípticas massivas são as mais afetadas pelo enriquecimento  $\alpha/Fe$ . As diferentes cores ilustram diferentes faixas de extinção: azul ( $A_V < -0.2$ ), verde ( $-0.2 \leq A_V \leq 0$ ) e vermelho ( $A_V > 0$ ). Podemos perceber que as galáxias com extinção mais negativa são as que possuem os maiores resíduos.

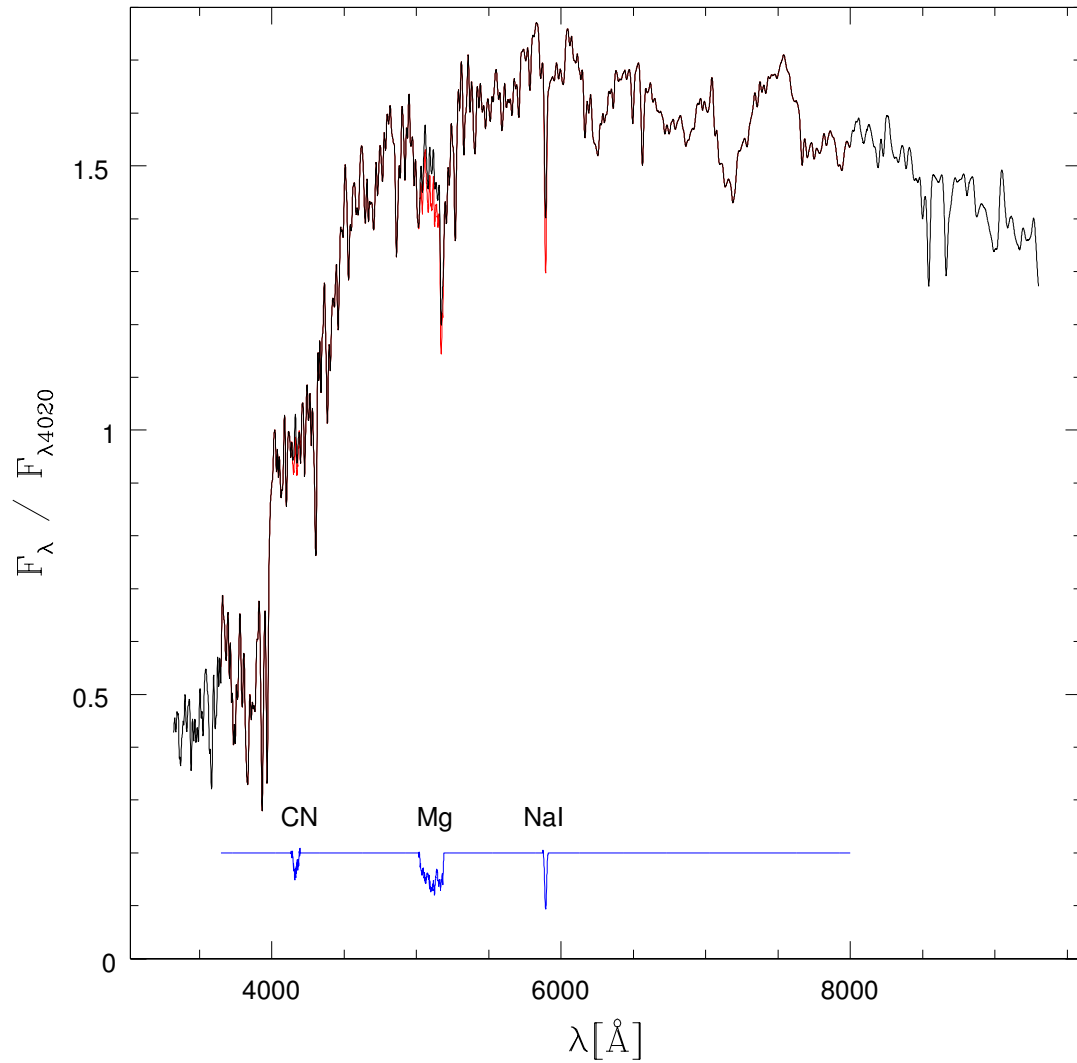


Fig. 4.18: Galáxia teste com acréscimo de resíduo médio nas bandas CN (4132–4196  $\text{\AA}$ ), Mg (5015–5190  $\text{\AA}$ ) e Na (5874–5911  $\text{\AA}$ ) simulando um enriquecimento de elementos  $\alpha$  em comparação à abundância solar.

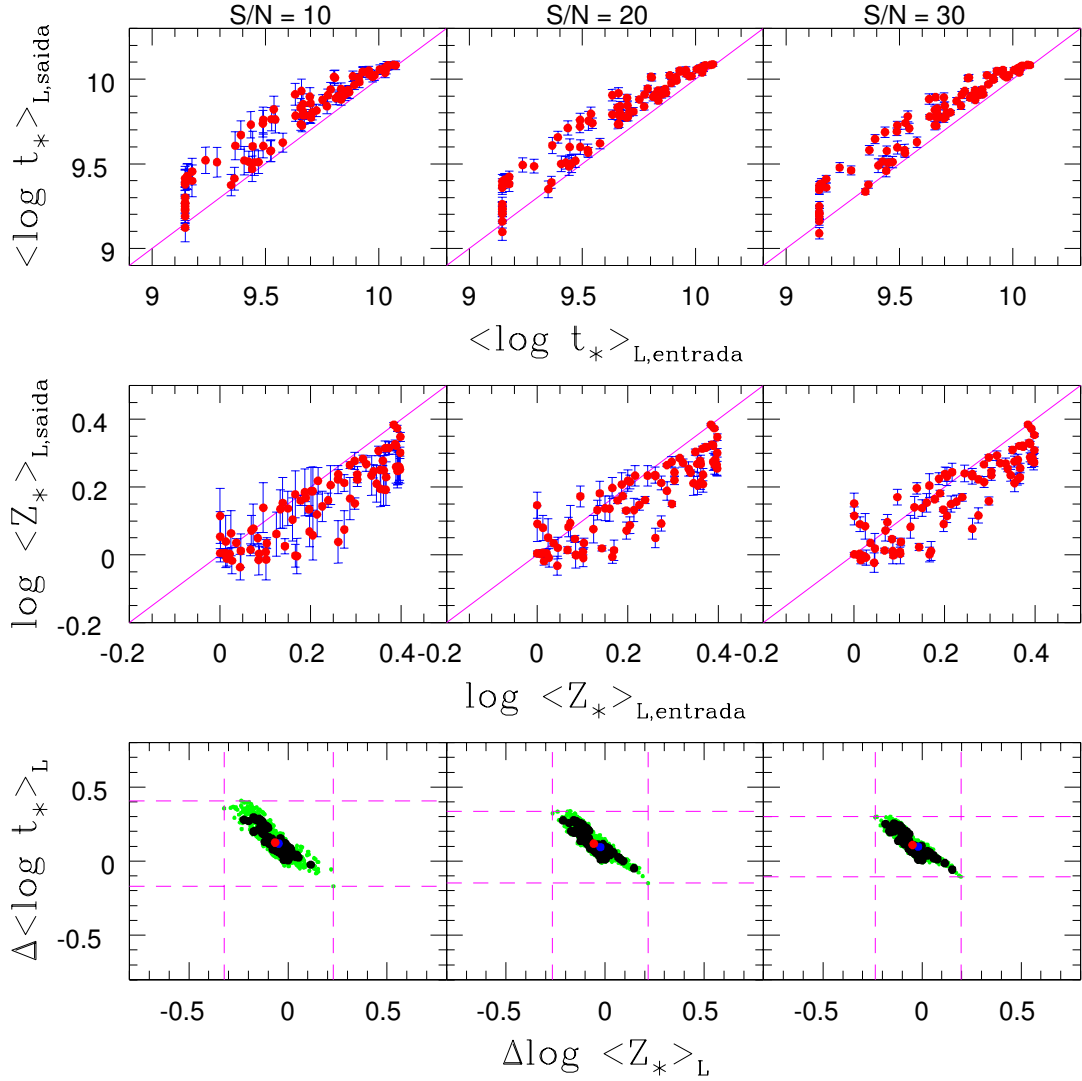


Fig. 4.19: Idade média e metalicidade média de entrada versus saída e degenerescência entre idade e metalicidade para 85 galáxias teste geradas no *quadrante superior direito* do diagrama  $\langle \log t_* \rangle_L$  versus  $\log \langle Z_* \rangle_L$  ( $0 \leq \log \langle Z_* \rangle_L \leq 0.4$  e  $8.9 \leq \langle \log t_* \rangle_L \leq 10.2$ ) **com acréscimo de um espectro de resíduo médio nas bandas CN (4132–4196 Å), Mg (5015–5190 Å) e Na (5874–5911 Å)**. Cada galáxia teste enriquecida por elementos  $\alpha$  foi perturbada 10 vezes para  $S/N = 10, 20$  e  $30$  em torno de  $\lambda_0 = 4020$  Å (utilizou-se um espectro de erro característico da SDSS - figura 2.7).

*Painéis superiores: Idade estelar média ou metalicidade estelar média de entrada versus saída:* Pontos são as médias das 10 versões perturbadas com as suas respectivas barras de erro. A linha cheia mostra onde deveriam cair os pontos idealmente, caso a síntese recuperasse com perfeição os valores de entrada.

*Painel inferior: Diagrama de degenerescência entre idade estelar média e metalicidade estelar média:* Pontos em verde são as 850 versões perturbadas das galáxias teste e pontos em preto são as médias das 10 versões perturbadas de cada galáxia.



Conclui-se que a introdução de resíduos em determinadas bandas e de um vínculo físico que não proíbe soluções com  $A_V < 0$ , ajuda a aumentar os erros no sentido de superestimar as idades médias.

Assim, a ocorrência de  $A_V < 0$  nestas simulações mostra que linhas mais profundas são melhor modeladas aumentando a idade, o que torna o espectro mais vermelho e acarreta  $A_V < 0$ . A magnitude do efeito ( $A_V \sim -0.1$ ) é muito semelhante com a observada para a amostra de galáxias reais. A metalicidade é subestimada devido à relação de degenerescência entre idade e metalicidade (anti-correlação), como a idade aumenta muito para ajustarmos as linhas, existe um efeito em sentido contrário que diminui a metalicidade.

#### 4.7.2 Configuração 2 ( $A_V \geq 0$ )

Aqui apresentamos o mesmo conjunto de simulações realizados anteriormente, contudo, impomos o vínculo físico  $A_V \geq 0$ . Na figura 4.20, mostramos para  $S/N = 10, 20$  e  $30$  os valores de entrada contra saída da idade estelar média e metalicidade estelar média e a degenerescência no plano  $\Delta\langle\log t_\star\rangle_L \times \Delta\log\langle Z_\star\rangle_L$ .

Apesar de não existir mais superestimação de idades e subestimação de metalicidades, ainda vemos deslocamentos em relação aos valores de entrada no diagrama  $\langle\log t_\star\rangle_L \times \log\langle Z_\star\rangle_L$ . Estes não são sistemáticos como na configuração 1, porém, a introdução de um resíduo médio nas bandas CN (4132–4196 Å), Mg (5015–5190 Å) e Na (5874–5911 Å) continua provocando deslocamentos não negligenciáveis neste diagrama. Os erros vão desde  $-0.19 < \Delta\langle\log t_\star\rangle_L < 0.28$  dex para a idade estelar média e para a metalicidade estelar média de  $-0.20 < \Delta\log\langle Z_\star\rangle_L < 0.28$  dex para  $S/N \geq 10$ .

Apesar de não existir uma tendência monótona como antes, temos ainda uma “preferência” por idades mais velhas e metalicidades *maiores* do que a dos valores de entrada. Nessa configuração todas as galáxias apresentaram extinção nula ( $\sim$  mesmo valor de entrada).

#### 4.7.3 Configuração 3 ( $A_V \geq 0$ e elementos $\alpha$ mascarados)

Nessa configuração, procuramos mostrar que ao mascararmos as regiões enriquecidas por elementos  $\alpha$ , recuperamos os parâmetros de entrada com mesma precisão da simulação sem acréscimo de resíduo.

Para a configuração 3, temos os resultados na figura 4.21 de valores de entrada contra saída para idade estelar média e metalicidade estelar média e gráfico de degenerescência entre idade e metalicidade. Os erros para essa última configuração vão desde  $-0.18 < \Delta\langle\log t_\star\rangle_L < 0.23$  dex para a idade estelar média e para a metalicidade estelar média de  $-0.16 < \Delta\log\langle Z_\star\rangle_L < 0.15$  dex para  $S/N \geq 10$  (figura 4.21 para  $S/N = 10, 20$  e  $30$  respectivamente).

#### 4.7.4 Sobre as Simulações: Configuração 1, 2 & 3

Podemos ver que a introdução de um resíduo sistemático nas bandas Mg, Na e CN faz com que incorramos em erros não negligenciáveis que não podem ser minimizados devido a termos uma deficiência em nossa base, isto é, estamos utilizando uma base

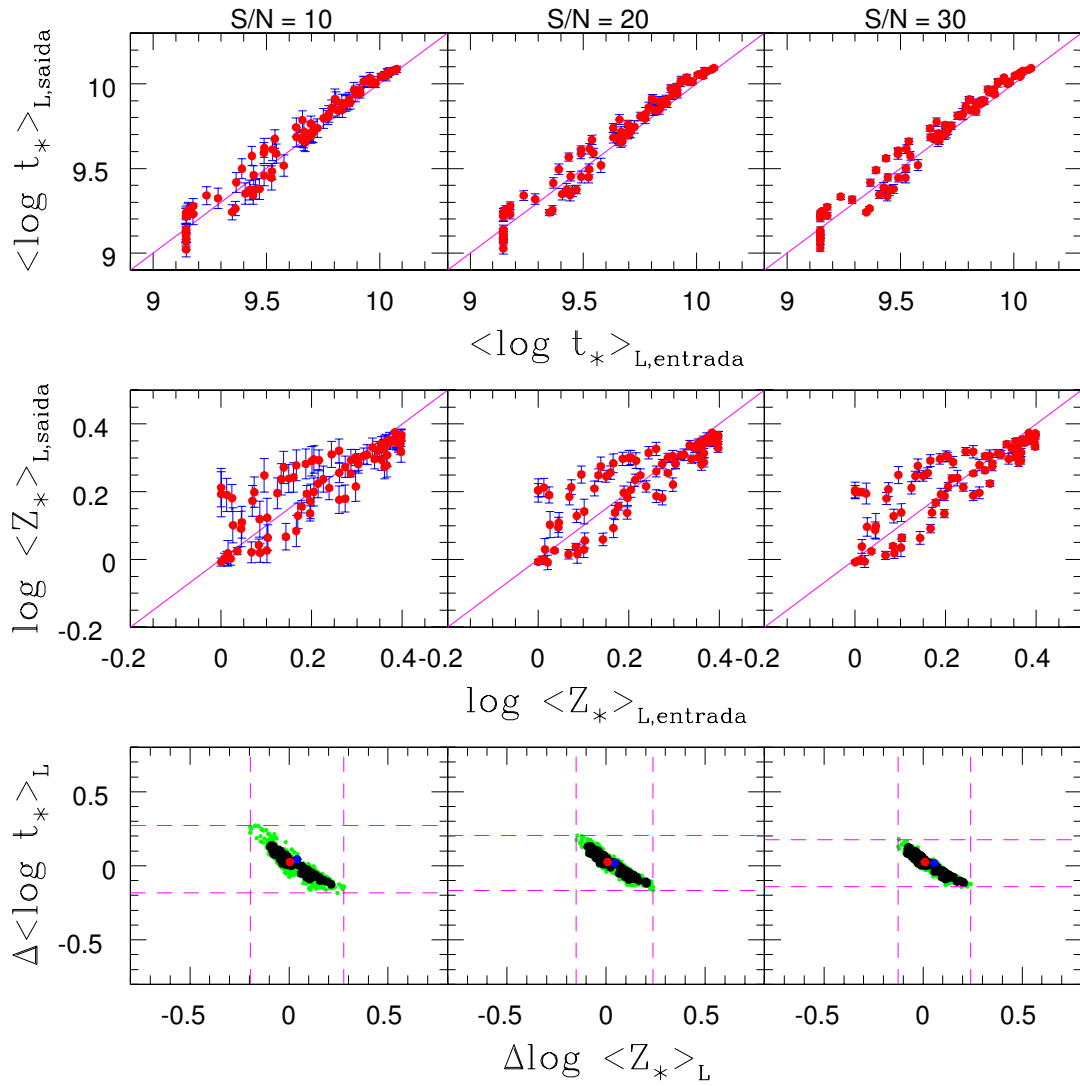


Fig. 4.20: A diferença dessa configuração para a anterior (figura 4.19) é a adição do vínculo físico  $A_V \geq 0$ .

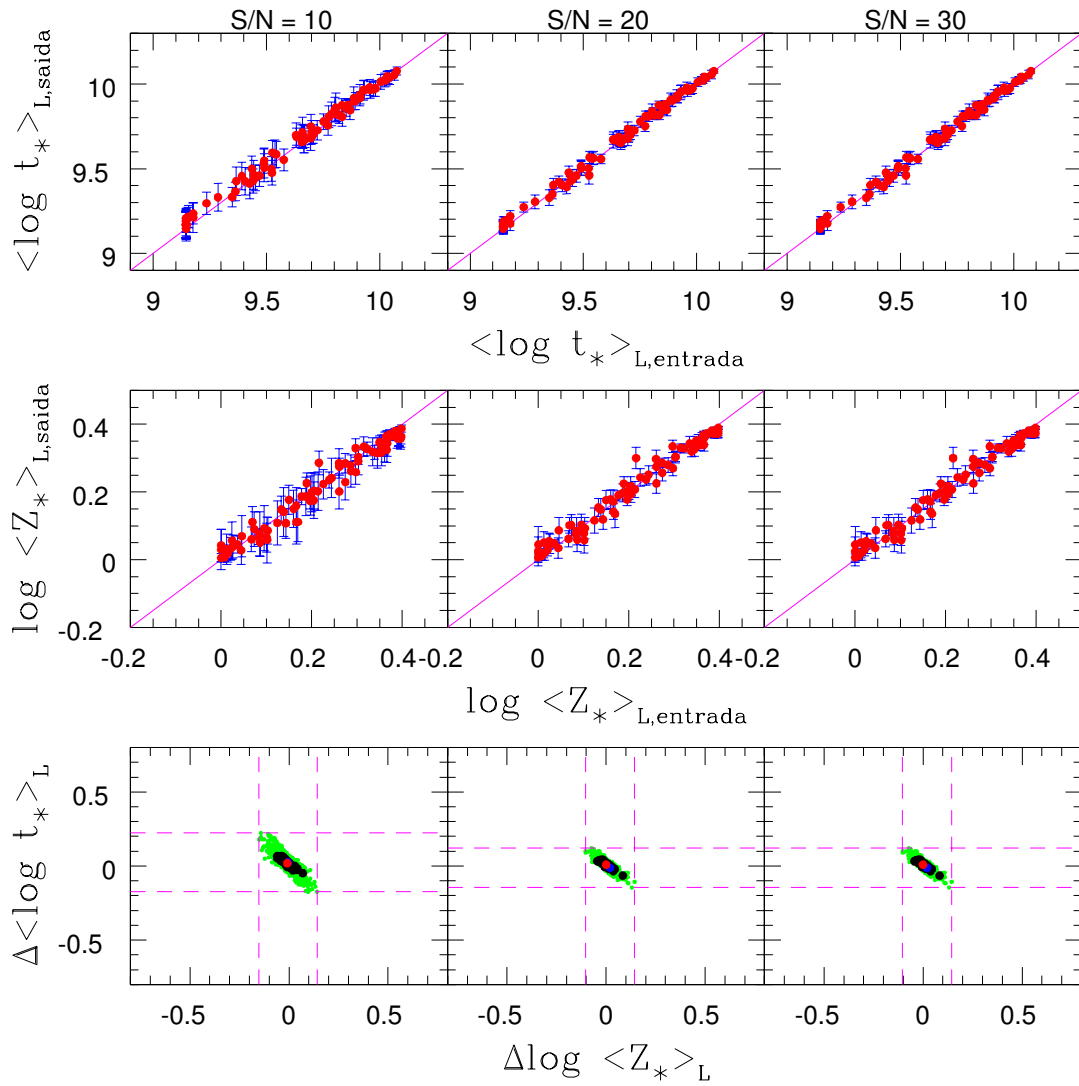


Fig. 4.21: Como as figuras 4.19 e 4.20, mas para a configuração 3, isto é,  $A_V \geq 0$  e mascarando os elementos  $\alpha$ .

escalonada com mistura solar de elementos químicos. Estes erros são da ordem de 0.2 dex para  $S/N \geq 10$  em  $\Delta \langle \log t_{\star} \rangle_L$  e  $\Delta \log \langle Z_{\star} \rangle_L$ . Outro ponto importante é em relação a extinções negativas. Se a permitirmos (configuração 1, subseção 4.7.3) causamos efeitos sistemáticos na sobrestimação das idades médias e subestimação das metalicidades, indicando que devemos impor um vínculo físico de positividade nas extinções ( $A_V \geq 0$ ).

As simulações tiveram um papel fundamental na determinação das magnitudes das discrepâncias observadas para a amostra de galáxias da SDSS incluindo ou não abundâncias sobre solares de elementos  $\alpha$ . Vimos que para idades altas e metalicidades altas os valores para o vetor de população não são confiáveis, e assim não podemos determinar com precisão a história de formação estelar e enriquecimento químico dessas galáxias.

#### 4.8 Possíveis cenários para enriquecimento $\alpha/Fe$

As variações encontradas nas abundâncias de  $\alpha/Fe$  mostram que nossas bases (N, O, P e Q) de SSP não modelam corretamente galáxias elípticas massivas. As populações estelares computadas por BC03 são escalonadas em relação a abundâncias solares, portanto é natural que uma galáxia com enriquecimento  $\alpha$  apresente resíduos quando compara-se  $O_{\lambda}$  com  $M_{\lambda}$ .

No trabalho de BC03, na análise de índices espectrais, chega-se à mesma conclusão. SSP's velhas não conseguem reproduzir bem índices de Lick como  $M_{g1}$ ,  $M_{g2}$ , entre outros para galáxias com alta dispersão de velocidade, mostrando existir uma abundância relativa diferente de elementos pesados entre galáxias massivas e estrelas galácticas utilizadas para os modelos (ver figuras 14 e 15 desse trabalho).

Existem três cenários possíveis (descritos sucintamente em Faber et al. 1992) para a origem ao excesso  $\alpha/Fe$ . Todos eles utilizam o fato de elementos leves serem produzidos em supernova tipo II, enquanto que o Fe é produzido predominantemente em supenovas tipo Ia.

1. **Taxas de formação estelar diferentes:** Teríamos que formar estrelas de uma maneira muito mais rápida e eficiente em galáxias elípticas gigantes para termos uma abundância maior de elementos leves em detrimento do Fe (resultado da figura 4.17).

Se admitirmos que há uma taxa diferenciada de formação estelar, consegue-se explicar a idéia acima, pois estrelas massivas que explodem em supernovas do tipo II vivem cerca de  $10^8$  anos (estas não possuem companheiras), ao passo que supernovas do tipo Ia estão em sistemas binários onde ocorre transferência de matéria, sendo um processo mais lento, que demora  $\sim 1$  Gano. Existe um retardo entre a formação de elementos leves e Fe, causando assim um enriquecimento de elementos  $\alpha$  para a primeira geração de estrelas formadas. Se toda formação estelar se concentra em  $< 1 \times 10^9$  anos, isso implica que estrelas de segunda “geração” têm  $\alpha$ , mas não Fe. Em suma,  $\alpha/Fe$  seria devido a uma fase rápida de formação estelar. Esse é o cenário mais aceito para explicar  $\alpha/Fe$ .

2. **Efeitos de seleção de elementos:** Ventos galácticos podem fazer com que o material processado em uma estrela seja perdido pela galáxia, não enriquecendo o meio interestelar. Assim sendo, para conseguirmos explicar as observações (figura 4.17), galáxias elípticas gigantes têm de perder mais Fe do que elementos  $\alpha$ , ou elípticas compactas têm de perder mais elementos  $\alpha$  do que Fe. A segunda hipótese parece ser mais provável já que galáxias muito massivas possuem bastante energia potencial gravitacional e, então, seria mais difícil exceder a velocidade de escape. Entretanto, os modelos de perdas de elementos por ventos galácticos sempre mostram um excesso de perda de Fe em relação aos outros (Arimoto & Yoshii 1987), não podendo explicar a razão supersolar  $\alpha/Fe$ .
3. **Função de massa inicial variável:** Supondo que quanto mais massiva uma galáxia elíptica é, mais a sua função de massa varia no sentido de privilegiar a existência de estrelas massivas em comparação ao restante de estrelas, concluímos que mais supernovas do tipo II seriam formadas. Teríamos mais produção de elementos leves em detrimento do Fe em elípticas gigantes do que compactas. Alguns mecanismos são propostos para o nascimento dessas estrelas como: choques dos braços espirais, compressão de nuvem de gás devido a expansão de regiões HII e remanescentes de supernovas bem como compressões da nuvem de gás molecular na interface entre duas dessas. Esse último parece ser o modo dominante em elípticas (que não possuem discos nem braços espirais). A pressão de uma nuvem de gás  $P$  é proporcional a densidade do gás  $\rho$ , e  $\rho \propto$  dispersão de velocidade do gás ( $\sigma_{\text{gás}}$ ), portanto quanto maior  $\sigma_{\text{gás}}$  maior a criação de estrelas massivas. Este cenário parece favorável pois as galáxias mais massivas possuem maior dispersão de velocidade e isto explicaria o surgimento de uma razão super solar de  $\alpha/Fe$  que aumenta com a dispersão de velocidade estelar.

#### 4.9 O Problema de Síntese para Galáxias Elípticas

Assim, mostramos que o problema de síntese de populações estelares para galáxias massivas é mais complexo do que se imagina. A conclusão de que existe uma abundância de elementos  $\alpha$  em relação ao Fe diferente da solar em galáxias elípticas nos leva a uma possível solução para este problema, utilizar uma base de SSPs computada para abundâncias de elementos não escalonadas com a solar.

Podemos também dizer que não se pode confiar nos resultados para estas galáxias, pois as simulações mostram que existem erros não negligenciáveis quando temos uma razão  $\alpha/Fe$  diferente da solar nos valores obtidos pela síntese (por exemplo: idade média e metalicidade média). Conseqüentemente, a tentativa de derivarmos a história de evolução estelar e enriquecimento químico de elípticas fica interrompida até que possamos utilizar uma nova base que leve em conta estes efeitos.

# Capítulo 5

## Considerações Finais e Perspectivas Futuras

*“Um pouco de técnica aqui e ali  
Também pode divertir o pensador  
Penso, audaz, que em tal campo  
Juntos faremos algo de bom”*

**Albert Einstein**  
(Ditado Traduzido)

Ao longo desse trabalho, procuramos desenvolver uma abordagem baseada em testes que revelaram o real potencial de nosso método de síntese. Através de simulações com galáxias teste criadas, sabendo-se seus parâmetros de entrada e comparando-os com os parâmetros de saída obtidos pelo código **STARLIGHT03**, verificamos que o método produz resultados bastante satisfatórios. Outra forma de verificarmos o nível de segurança do método foi aplicar o mesmo a uma amostra de 50362 galáxias reais da SDSS, realizando um teste empírico. Ao compararmos nossos resultados com resultados obtidos de maneiras distintas, conseguimos desse modo um outro indicativo de confiabilidade de nosso código que não possuíam conexão direto com a nossa síntese espectral.

Através das primeiras simulações realizadas, conseguimos recuperar os parâmetros com precisão excelente. Os resíduos de saída para o vetor de população condensado em componentes jovem ( $x_Y$ ), de idade intermediária ( $x_I$ ) e velha ( $x_O$ ), possuem incertezas menores que  $\Delta x_Y = 5\%$ ,  $\Delta x_I = 10\%$  e  $\Delta x_O = 10\%$  para  $S/N \geq 10$ . As incertezas para a extinção são menores que  $\Delta A_V \leq 0.05$  mag, para a massa estelar são  $\Delta \log M_\star \leq 0.1$  dex e para a dispersão de velocidade são  $\Delta \sigma_\star \leq 12$  km s<sup>-1</sup> quando o  $S/N \geq 10$  por pixel.

Em relação à aplicação do código a galáxias da SDSS, além de conseguirmos ajustar de forma excelente os espectros observados (figuras 2.1 e 2.2), obtemos resultados que estão de acordo com medidas realizadas de maneiras independentes, como por exemplo: a extinção nebulosa versus extinção estelar, idade estelar média versus largura equivalente em  $H\alpha$  e  $D_n4000$  e a massa estelar versus dispersão de velocidade (figuras 3.2, 3.3 e 3.4).

Apesar de todo esse sucesso em relação ao método, uma análise cuidadosa dos dados nos indica dois possíveis “problemas” (capítulo 4). Esses problemas poderiam ser relacionados com o método e/ou base utilizada. O primeiro refere-se a um duplo pico de formação estelar, sendo um deles em uma idade média de  $\sim 10^{9.2}$  anos o que indicaria que existiu um surto de nascimento de estrelas nesse período, porém não há registro algum. Para o segundo problema, vemos que existe uma porção do diagrama idade estelar média versus metalicidade estelar média para altas idades e altas metalicidades na qual não existem galáxias, ou seja, um limite superior que podemos ver por uma linha clara nesse diagrama, que como consequência indicaria que as galáxias mais velhas não são as mais metálicas e vice-versa, um resultado de peso muito forte na astrofísica.

Como primeiro reflexo, fizemos testes mudando elementos de nossa  $\mathcal{B}_N$ , que até esse momento era utilizada, criando assim as bases O, P e Q. Estas mudanças foram necessárias para investigarmos se estes problemas eram devido a uma limitação da base. Mostramos que o primeiro desses problemas era facilmente solucionado ao introduzirmos elementos com metalicidades muito baixas  $Z = 0.005, 0.02 Z_{\odot}$ . Porém, o segundo efeito continuava a existir com todas as novas bases.

Uma análise mais cuidadosa dos espectros mostrou que as galáxias que se situavam na estranha linha que limita a porção superior direita do diagrama idade estelar média versus metalicidade estelar média eram elípticas, e estas possuíam resíduos sistemáticos em determinadas bandas (Mg, Na e CN) que aumentam à medida que suas massas aumentam (figura 4.17). Esse efeito é devido a um enriquecimento de elementos  $\alpha$ . Além disto, galáxias elípticas sintetizadas possuem extinções levemente negativas, que eram permitidas por motivos estatísticos.

Para melhor testarmos os efeitos de degenerescência do código, realizamos uma nova série de simulações que visavam investigar se uma dada galáxia com idades altas e metalicidades altas era deslocada nesse diagrama. Estas galáxias eram deslocadas dentro do limite imposto pela relação de degenerescência entre idade e metalicidade. Isto implica que os resultados são bastante confiáveis dentro de uma precisão de até  $\sim 0.2$  dex. Concluímos que a não existência de galáxias com alta idade e metalicidade não é devida a problemas com o código.

Após estas simulações terem sido realizadas, fizemos outros testes com as mesmas galáxias, porém agora com a introdução nas mesmas de um resíduo médio em certas bandas de elementos  $\alpha$  (Mg, Na, CN). Verificou-se que isto gerava no processo de síntese efeitos não negligenciáveis, como o deslocamento de algumas galáxias no diagrama  $\langle \log t_{\star} \rangle_L$  versus  $\log \langle Z_{\star} \rangle_L$  da ordem de 0.4 dex. Apesar disto, os ajustes dos espectros ainda podem ser considerados bons. Outro ponto importante que também foi notado é o fato das extinções saírem levemente negativas ( $A_V < 0$ ). Linhas mais profundas são melhor modeladas aumentando a idade das componentes da base, que têm espectros mais vermelhos, acarretando em  $A_V < 0$ . A magnitude do efeito ( $A_V \sim -0.1$ ), é muito parecida com a “observada” nas galáxias da SDSS.

## 5.1 Perspectivas

Apesar dos problemas discutidos nessa dissertação, nosso método de síntese espectral é muito poderoso, e, em conjunto com super-bases-de-dados permite abordar

uma longa lista de questões científicas interessantes. Como os problemas mais graves detectados dizem respeito a galáxias elípticas (velhas, sem formação estelar, massivas), é estrategicamente melhor começar estudando as que não têm problemas. Galáxias com formação estelar (em geral, espirais e irregulares) são livres de problemas, e têm a vantagem de que fornecem informações valiosas através de suas linhas de emissão. Neste momento estamos iniciando dois estudos:

- (1) Diz respeito à dependência das extinções estelares e nebulares com parâmetros físicos como idade, metalicidade estelar, metalicidade nebular, etc.
- (2) Utilizando uma diferente metodologia e aproveitando a informação sobre metalicidade estelar fornecida pela síntese seguindo a linha do trabalho de Tremonti et al. (2004), obteremos a relação entre massa e metalicidade de galáxias da SDSS.

Em relação a galáxias elípticas, que ocuparam a maior parte desse trabalho, a perspectiva futura é implementarmos SSPs com enriquecimento  $\alpha$  na base. Utilizaremos, então, nosso método de síntese para tentarmos recuperar a história de formação estelar e enriquecimento químico dessas galáxias. Existem dois trabalhos recentes que calculam espectros estelares com enriquecimento  $\alpha$ : Munari et al. (2005) e Coelho et al. (2005). Quando isto for incluído em programas de síntese evolutiva (exemplo, o GALAXEV, de Bruzual & Charlot), poderemos gerar espectros de SSPs com enriquecimento  $\alpha$ , e colocá-los na base. Planejamos fazer isto assim que os novos modelos forem disponibilizados.

Por fim, em relação à melhoria do nosso código, estamos implementando várias novas técnicas de amostragem para o apressarmos. Algumas dessas novas técnicas, como o monitoramento do “Burn-in”, estão explicadas de forma mais detalhada no apêndice B.



## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Abazajian K. et al., 2003, AJ, 126, 2081.
- [2] Abazajian K. et al., 2004, AJ, 128, 502.
- [3] Arimoto N., Yoshii Y., 1987, A&A, 173, 23.
- [4] Baade W., 1944, ApJ, 100, 147.
- [5] 1986 A&A...166...83B 1.000 09/1986 A F G R C S N U Bica, E.; Alloin, D.
- [6] Bica E., 1988, A&A, 195, 76.
- [7] Bica E., Alloin D., Schmitt H. R., 1994, A&A, 283, 805.
- [8] Bruzual G., Charlot S., 2003, MNRAS, 344, 1000 (BC03).
- [9] Calzetti D., Kinney A. L., Storchi-Bergmann T., 1994, ApJ, 345, 245.
- [10] Chabrier G., 2003, PASP, 115, 763.
- [11] Coelho P., Barbuy B., Melendez J., Schiavon R., Castilho B., 2005, astro-ph/0505511.
- [12] Cid Fernandes R., Sodr e L., Schmitt H. R., Le o J. R. S., 2001, MNRAS, 325, 60.
- [13] Cid Fernandes R., Mateus A., Sodr e L., Stasi nska G., Gomes J. M., 2005, MNRAS, 358, 363.
- [14] Cid Fernandes R., Gu Q., Melnick K., Terlevich E., Terlevich R., Kunth D., Rodrigues lacerda R., Joguet B., 2004, MNRAS, 355, 273, (CF04).
- [15] Dottori H. A., 1981, Ap&SS, 80, 267.
- [16] Faber S. M., 1972, A&A, 20, 361.
- [17] Faber S. M., Worthey G., Gonzalez J. J., 1992, IAU, 255, 149.
- [18] Garcia-Rissmann A., Vega L. R., Asari N. V., Cid Fernandes R., Gonz alez Delgado R. M., Storch-Bergmann T., 2005, MNRAS, 359, 765.
- [19] Heavens A. F., Jimenez R., Lahav O., 2000, MNRAS, 317, 965.
- [20] Kauffmann G. et al., 2003, MNRAS, 346, 1055 (K03a).

- 
- [21] Kauffmann G. et al., 2003, MNRAS, 341, 54 (K03b).
- [22] Kauffmann G. et al., 2003, MNRAS, 341, 33 (K03c).
- [23] Kennicutt R. C. Jr., 1998, ARA&A, 36, 189.
- [24] Le Borgne J. -F. et al., 2003, A&A, 402, 433.
- [25] Maraston C., Greggio L., Renzini A., Ortolani S., Saglia R. P., Puzia T. H., Kissler-Patig M., 2003, A&A, 400, 823.
- [26] Morgan W. W., 1956, PASP, 68, 509.
- [27] Munari U., Sordo R., Castelli F., Zwitter T., 2005, astro-ph/0502047.
- [28] Panter B., Heavens A. F., Jimenez R., 2004, MNRAS, 355, 764.
- [29] Pelat D., 1997, MNRAS, 284, 365.
- [30] Pelat D., 1998, MNRAS, 299, 877.
- [31] Salasnich B., Girardi L., Weiss A., Chiosi C., 2000, A&A, 361, 1023.
- [32] Schmidt A. A., Copetti M. V. F., Alloin D., Jablonka P., 1991, MNRAS, 249, 766.
- [33] Spinrad H., Tinsley B. M., 1972, ApJ, 171, 397.
- [34] Strauss M. A. et al., 2002, AJ, 124, 1810.
- [35] Tinsley B. M., 1968, ApJ, 151, 547.
- [36] Thomas D., Maraston C., 2003, A&A, 401, 429.
- [37] Thomas D., Maraston C., Bender R., 2003, MNRAS, 339, 897.
- [38] Trager S. C., Worthey Guy, Faber S. M., Burstein David, Gonzalez J. Jesus, 1998, ApJS, 116 , 1.
- [39] Tremonti C., 2003, PhD thesis, Johns Hopkins Univ.
- [40] Vega L. R., 2004, Dissertação de Mestrado, Universidade Federal de Santa Catarina.
- [41] Wood D. B., 1966, ApJ, 145, 36.
- [42] Worthey G., Faber S. M., Gonzalez J. J., 1992, ApJ , 398, 69.
- [43] Worthey G., 1994, ApJS, 94, 687.

# Apêndice A

## Nomenclatura

---

$\alpha$	Elementos formados por núcleos de He como: O, Ne, Mg, Si, S, Ar, Ca, Ti mais dois elementos: N e Na
$\Delta_\lambda$	Desvio percentual entre espectro modelado e observado
$\Delta\langle\log t_\star\rangle_L$	Diferença entre idade média estelar de saída e de entrada
$\Delta\log\langle Z_\star\rangle_L$	Diferença entre metalicidade média estelar de saída e de entrada
$\lambda_0$	Comprimento de onda de normalização
$\sigma_\star$	Dispersão de velocidade
$\zeta$	Espaço de parâmetros
$A_V$	Extinção na banda V
BASES N, O, P e Q	São denotadas por $\mathcal{B}_N$ , $\mathcal{B}_O$ , $\mathcal{B}_P$ e $\mathcal{B}_Q$
BC03	Bruzual & Charlot (2003)
CF04	Cid Fernandes et al. 2004b
EW	Largura Equivalente
IMF	“Initial mass function” (função de massa inicial). É o número de estrelas formadas entre massa $m$ e $m + dm$ em um tempo $t$ .
ISM	“Interstellar Medium” (meio interestelar)
$\langle\log t_\star\rangle$	Idade estelar média definida como $\sum_i^{N_\star} \log t_i \times x_i$
$\log\langle Z_\star\rangle$	Metalicidade estelar média definida como $\log(\sum_i^{N_\star} Z_i \times x_i)$
$M_\lambda$	Espectro modelado
$M_\star$	Massa estelar
MOPED	“Multiple Optimized Parameter Estimation and Data Compression”
$N_\star$	Número de populações estelares simples na base
$O_\lambda$	Espectro observado
SDSS	“Sloan Digital Sky Survey”
SPEE	Síntese de população estelar evolutiva
SPESE	Síntese de população estelar semi-empírica
SSP	“Simple stellar population” (população estelar simples)
STARLIGHT	Código de síntese utilizado extensivamente nesse trabalho
$S/N$	Relação sinal/ruído
$t_i$	Idade da população estelar $i$
$\vec{x}$	Vetor de populações. Composto pelas frações de luz de cada elemento da base $x_i$ em relação ao fluxo total.
$Z_i$	Metalicidade da população estelar $i$

---

# Apêndice B

## Uma Formulação Probabilística para a Síntese de Populações Estelares

Procuramos mostrar nesse apêndice uma formulação matemática mais rigorosa em relação ao problema de síntese. O objetivo é complementarmos o conhecimento desenvolvido ao longo dessa dissertação para aqueles que desejam um tratamento mais rigoroso da estatística utilizada.

Essa estatística serve para encontrarmos os valores para os parâmetros da síntese, ou melhor, a distribuição de probabilidades para os parâmetros (posteriores) utilizando a técnica de amostragem explicada a seguir, apesar de não a utilizarmos diretamente nesse trabalho mas que pode ser facilmente implementada no futuro. Outra novidade que está sendo desenvolvida é o monitoramento da convergência a fim de apressarmos o código. Já existe uma nova versão que faz uso desses novos tópicos aqui mencionados, porém ainda em fase de teste.

### *B.1 Estatística Bayesiana*

Quando estamos trabalhando com um certo conjunto de dados, estes não nos dizem muito em relação às propriedades que estão a ser observadas. Assim, precisamos processar a informação a fim de reconstruirmos a informação incompleta que nos é creditada. Portanto, precisamos de um esquema matemático para fazermos a análise dos dados.

Utilizaremos a teoria de probabilidades para fazermos as inferências necessárias em relação à informação incompleta dos “dados”.

Probabilidades são sempre números entre  $[0,1]$  e obedecem a duas regras fundamentais:

$$\text{prob}(A) + \text{prob}(B) = 1 \tag{B.1}$$

onde, A e B são determinados eventos. A primeira equação é chamada de regra da SOMA e a posterior, regra do PRODUTO:

$$\text{prob}(AB) = \text{prob}(A|B)\text{prob}(B) \tag{B.2}$$

Onde,  $prob(A)$  é a probabilidade associada a um dado estado  $A$ ;  $prob(AB)$  é a probabilidade de  $A$  e  $B$  ocorrerem e  $prob(A|B)$  é a chamada probabilidade condicional, ou seja, é a probabilidade de  $A$  ocorrer dado que  $B$  tenha ocorrido.

Um dos importantes corolários que podem ser derivados a partir dessas duas relações é o teorema de Bayes. Deduziremos o teorema a seguir:

$$prob(AB) = prob(BA) \Rightarrow prob(A|B)prob(B) = prob(B|A)prob(A) \quad (B.3)$$

Portanto, chega-se ao teorema de Bayes:

$$prob(B|A) = \frac{prob(A|B)prob(B)}{prob(A)} \quad (B.4)$$

Note que as probabilidades condicionais  $prob(A|B)$  e  $prob(B|A)$  são diferentes. Se tivermos um dado modelo  $M$  com alguns parâmetros  $\aleph$  e quisermos ajustar a um conjunto de dados  $D$ , poderemos utilizar o teorema de Bayes assumindo-se que sabemos o suficiente acerca do modelo dado. Se tivermos o valor de um parâmetro, podemos calcular a probabilidade de qualquer conjunto de dados para que seja gerado pelo modelo. A isto damos o nome de “likelyhood” (verossimilhança). Temos, então:

$$prob(\aleph|D, M) = \frac{prob(D|\aleph, M)prob(\aleph|M)}{prob(D|M)} \quad (B.5)$$

Ao resolvermos o problema, o que queremos obter é a probabilidade da distribuição dos parâmetros em relação a um conjunto de dados, lado esquerdo da equação anterior é chamado de posterior. À direita da equação temos a verossimilhança e o termo estranho  $prob(\aleph|M)$ , que é a probabilidade de distribuição dos parâmetros dado o modelo  $M$  e sem considerarmos o conjunto de dados, é chamado de prior. O denominador é o termo de marginalização que, para o presente trabalho, não será de muita importância.

Assumindo-se que os erros são gaussianos ( $\{\sigma_i\}$ ), que o conjunto de dados experimentais seja  $D = \{y_i, x_i, \sigma_i\}$ , que o modelo é  $M(x_i, \aleph)$  com parâmetros  $\aleph$  e  $N$  seja o número de dados que temos, a verossimilhança será dada por:

$$prob(D|\aleph, M) \sim e^{-\sum_{i=1}^N \frac{(y_i - M(x_i, \aleph))^2}{(2\sigma_i)^2}} = e^{-\frac{\chi^2}{2}} \quad (B.6)$$

Para o caso de um prior não informativo, ou seja, um prior uniforme, temos que o posterior será proporcional à verossimilhança.

$$prob(\aleph|D, M) \sim prob(D|\aleph, M) \quad (B.7)$$

Vemos que a “likelyhood” pode ser expressa como uma função do  $\chi^2$ . Daqui para frente iremos denominar energia  $E = \chi^2/2$ , isto é uma alusão à mecânica estatística e será útil para a discussão sobre como mapear o espaço de parâmetros de forma eficiente e obtermos os posteriores.

## B.2 Algoritmo de Metropolis

O grande problema agora é como mapear o espaço de parâmetros  $\zeta$  a fim de obtermos as distribuições dos posteriores. Poderíamos simplesmente pensar em gerar números randômicos e caminharmos no espaço  $\zeta$ . Porém, sem nenhum critério de amostragem, teríamos um posterior irreal sem sentido físico.

Portanto, é preciso de algum critério para mapearmos a região de maior probabilidade do espaço  $\zeta$ . Foi com esse intuito que foi criado o algoritmo de Metropolis. Esse algoritmo serve para fazer uma amostragem direta do espaço de parâmetros. Poderemos, então, obter as distribuições dos parâmetros em relação a um conjunto de dados, ou seja, os posteriores.

Podemos descrever o algoritmo de Metropolis com as seguintes etapas:

- Escolhemos um ponto inicial no espaço  $\zeta$ , ou seja um vetor de parâmetros inicial  $\aleph_0$ .
- Calculamos a verossimilhança  $prob(D|\aleph_0, M)$  de observar os dados experimentais dado os parâmetros  $\aleph_0$ .
- Obtemos um novo vetor de parâmetros amostrando a partir de uma “distribuição “proposta”  $q(\aleph_{i-1}, \aleph_i)$ .
- Compute a nova verossimilhança  $prob(D|\aleph_i, M)$ .
- Se  $prob(D|\aleph_i, M) > prob(D|\aleph_{i-1}, M)$ , então guarde o novo vetor de parâmetros  $\aleph_i$  como o novo ponto na cadeia. Isto é o equivalente a utilizarmos a noção de energia definida anteriormente, ou seja, se  $E_i < E_{i-1}$  então, guarda-se o novo valor dos parâmetros.
- Se  $prob(D|\aleph_i, M) < prob(D|\aleph_{i-1}, M)$ , ou o equivalente  $E_i > E_{i-1}$ , então gere uma variável randômica “ $u$ ” entre  $[0, 1]$ , definindo a probabilidade de transição entre o estado  $i-1$  para o estado final  $i$  como  $P_{(i-1) \rightarrow (i)} = \frac{prob(D|\aleph_i, M)}{prob(D|\aleph_{i-1}, M)}$ . Agora, se  $u < P_{(i-1) \rightarrow (i)}$ , aceite o novo conjunto de parâmetros  $\aleph_i$  e se  $u > P_{(i-1) \rightarrow (i)}$ , rejeite o novo estado.

Esse algoritmo é muito útil, pois ele tende a caminhar na direção de maior probabilidade no espaço de parâmetros, ou seja, em direção a regiões de menor energia, no entanto determinadas vezes a escolha do próximo passo se dá em direção a energias mais altas, assim não se prendendo a mínimos locais.

### B.2.1 “Burn-in”

No início, quando escolhemos um determinado ponto no espaço de parâmetros, quase nunca temos a sorte de já encontrarmos a região de maior probabilidade. Assim, enquanto o algoritmo de Metropolis vai sendo utilizado, os pontos iniciais dessa cadeia não constituem na realidade uma amostragem dos posteriores, a isto damos o nome de “burn-in”. Portanto, o “burn-in” são os pontos que devem ser descartados no procedimento de amostragem direta. Em princípio, é difícil dizer

quando uma determinada cadeia convergiu para a região de maior probabilidade. Podemos ter uma idéia do término do “burn-in” quando analisamos um gráfico de Energia versus iteração, mas esse será um resultado qualitativo.

É possível monitorarmos a convergência, usando esse tipo de algoritmo de Monte Carlo, com várias cadeias começando de pontos aleatórios no espaço  $\zeta$ . O teste de convergência de Gelman e Rubin monitora a variância de um parâmetro entre as várias cadeias. Iremos ilustrar o teste a seguir, considerando os últimos  $n$  pontos de  $m$  cadeias. Seja  $\psi_{ij}$  uma entrada para o vetor de parâmetros  $\aleph$  no ponto  $j$  na cadeia  $i$  com média  $\langle\psi_i\rangle$  para a cadeia  $i$  e média global, de todas as cadeias,  $\langle\psi\rangle$ . Podemos escrever a variância entre cadeias  $B$  e a variância intrínseca das cadeias  $W$ :

$$B = \frac{n}{m-1} \sum_{i=1}^m (\langle\psi_i\rangle - \langle\psi\rangle)^2 \quad (\text{B.8})$$

$$W = \frac{1}{m} \sum_{i=1}^m s_i^2 \quad (\text{B.9})$$

onde,  $s_i^2 = \frac{1}{n-1} \sum_{j=1}^n (\psi_{ij} - \langle\psi_i\rangle)^2$  e a quantidade

$$R = \frac{\frac{n-1}{n}W + \frac{1}{n}B}{W} \quad (\text{B.10})$$

deveria convergir para 1, porém um número aceitável para convergência é  $R < 1.2$  para todos os parâmetros. Assim, temos a indicação de que estaremos amostrando  $prob(\aleph|D, M)$ , e a partir do ponto onde isto ocorre, poderemos fazer a amostragem para a estimação dos parâmetros.

### B.3 Formalismo Aplicado à Síntese

Vimos, anteriormente, o desenvolvimento matemático necessário para um problema qualquer, na qual temos um dado modelo e gostaríamos de saber a distribuição probabilística dos parâmetros  $\aleph$ . Agora, iremos exemplificá-lo com o problema da síntese de populações estelares semi-empírica.

Por simples analogia com as equações obtidas anteriormente, podemos ver que:

- O modelo  $M$  depende exclusivamente do vetor de população  $\vec{x}$ , da extinção  $A_V$ , da velocidade da linha de visada centrada em  $v_*$  e da dispersão de velocidade  $\sigma_*$ . Portanto, os parâmetros para o problema da síntese são  $\aleph = \vec{x}, A_V, v_*, \sigma_*$ .
- O conjunto de dados  $D$  é justamente a faixa de espectro utilizada obtida através das observações  $O_\lambda$ . Os erros nas medidas  $\sigma_\lambda$  podem ser considerados como parte dos  $D$ .
- Claro que é preciso de um conjunto de hipóteses envolvidas: a representatividade da base, a lei de avermelhamento, os erros gaussianos, ... De acordo com as equações anteriores, podemos considerar as hipóteses fazendo parte do próprio modelo  $M$ , mas as designaremos por  $I$  ( $I$  = informações complementares ao modelo).

Assim, podemos facilmente reescrever o teorema de Bayes como:

$$P(\vec{x}, A_V, v_*, \sigma_* | D, \sigma_\lambda, I) = P(\vec{x}, A_V, v_*, \sigma_* | I) \frac{P(D | \vec{x}, A_V, v_*, \sigma_*, \sigma_\lambda, I)}{P(D | \sigma_\lambda, I)} \quad (\text{B.11})$$

Obtemos, então, uma maneira de avaliarmos a probabilidade de  $\vec{x}, A_V, v_*, \sigma_*$  em função dos dados, dos erros observacionais e de outras informações plausíveis.

A verossimilhança é, como vimos anteriormente, uma forma de medir quão bom é o ajuste dos dados ao modelo. Assumindo os erros gaussianos, temos que a probabilidade dos dados dado os parâmetros é utilizando a analogia das equações anteriores:

$$P(D | \vec{x}, A_V, v_*, \sigma_*, \sigma_\lambda, I) \sim e^{-E} \quad (\text{B.12})$$

Onde a energia  $E$  dependerá de  $\vec{x}, A_V, v_*, \sigma_*$ , pois o modelo  $M_\lambda$  depende dessas quantidades como foi mostrado nas equações anteriores.

$$E(\vec{x}, A_V, v_*, \sigma_*) = \frac{1}{2} \chi^2 = \frac{1}{2} \sum_{\lambda} [(O_\lambda - M_\lambda) \omega_\lambda]^2 \quad (\text{B.13})$$

Vemos que  $O_\lambda$  é o fluxo do espectro observado no comprimento de onda  $\lambda$ ;  $M_\lambda$  é o fluxo sintético que vem do modelo utilizado e depende de  $x$  e  $A_V$ . Vemos que o peso  $\omega_\lambda$  é nada mais nada menos que o inverso do erro em  $O_\lambda$ , ou seja,  $\omega_\lambda = \frac{1}{\sigma_\lambda}$ .

As informações adicionais  $I$  incluem os vínculos que devem ser respeitados nos resultados da síntese. Estas informações estão presentes no prior e são dadas por:

- $\sum_{j=1}^{N_*} x_j = 1$ , ou seja, constitui um vínculo de normalização e limita a soma das componentes individuais a 1.
- $x_j \geq 0$ , que indica que as componentes individuais não podem ser menores que 0, ou seja, têm de ser positivas.
- Temos também a restrição da extinção, pois essa tem de ser positiva já que não há sentido físico para uma extinção negativa,  $A_V \geq 0$ . Em alguns casos, a fim de mapearmos a distribuição de probabilidades, retiramos esse vínculo para termos resultados estatísticos mais confiáveis (sem truncar a distribuição).

O código de síntese **STARLIGHT** utiliza como base os mesmos conceitos aqui desenvolvidos.



# Semi-empirical analysis of Sloan Digital Sky Survey galaxies – I. Spectral synthesis method

Roberto Cid Fernandes,<sup>1★</sup> Abílio Mateus,<sup>2★</sup> Laerte Sodré, Jr.,<sup>2★</sup> Grażyna Stasińska<sup>3★</sup>  
and Jean M. Gomes<sup>1★</sup>

<sup>1</sup>*Departamento de Física, CFM, Universidade Federal de Santa Catarina, PO Box 476, 88040-900 Florianópolis, SC, Brazil*

<sup>2</sup>*Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, São Paulo, SP, Brazil*

<sup>3</sup>*LUTH, Observatoire de Meudon, 92195 Meudon Cedex, France*

Accepted 2004 December 14. Received 2004 December 10; in original form 2004 October 18

## ABSTRACT

The study of stellar populations in galaxies is entering a new era with the availability of large and high-quality data bases of both observed galactic spectra and state-of-the-art evolutionary synthesis models. In this paper we investigate the power of spectral synthesis as a means to estimate the physical properties of galaxies. Spectral synthesis is nothing more than the decomposition of an observed spectrum in terms of a superposition of a base of simple stellar populations of various ages and metallicities, producing as output the star formation and chemical histories of a galaxy, its extinction and velocity dispersion. Our implementation of this method uses the recent models of Bruzual & Charlot and observed spectra in the 3650–8000 Å range. The reliability of this approach is studied by three different means: (1) simulations, (2) comparison with previous work based on a different technique, and (3) analysis of the consistency of results obtained for a sample of galaxies from the Sloan Digital Sky Survey (SDSS).

We find that spectral synthesis provides reliable physical parameters as long as one does not attempt a very detailed description of the star formation and chemical histories. Robust and physically interesting parameters are obtained by combining the (individually uncertain) strengths of each simple stellar population in the base. In particular, we show that, besides providing excellent fits to observed galaxy spectra, this method is able to recover useful information on the distributions of stellar ages and, more importantly, stellar metallicities. Stellar masses, velocity dispersion and extinction are also found to be accurately retrieved for realistic signal-to-noise ratios.

We apply this synthesis method to a volume-limited sample of 50 362 galaxies from the SDSS Data Release 2, producing a catalogue of stellar population properties. Emission lines are also studied, their measurement being performed after subtracting the computed starlight spectrum from the observed one. A comparison with recent estimates of both observed and physical properties of these galaxies obtained by other groups shows good qualitative and quantitative agreement, despite substantial differences in the methods of analysis. The confidence in the present method is further strengthened by several empirical and astrophysically reasonable correlations between synthesis results and independent quantities. For instance, we report the existence of strong correlations between stellar and nebular metallicities, stellar and nebular extinctions, mean stellar age and equivalent width of H $\alpha$  and 4000-Å break, and between stellar mass and velocity dispersion.

**Key words:** galaxies: evolution – galaxies: fundamental parameters – galaxies: statistics – galaxies: stellar content.

## 1 INTRODUCTION

Galaxy spectra encode information on the age and metallicity distributions of the constituent stars, which in turn reflect the star formation and chemical histories of the galaxies. Retrieving this

\*E-mail: cid@astro.ufsc.br (RCF); abilio@astro.iag.usp.br (AM);  
laerte@astro.iag.usp.br (LS); grazyna.stasinska@obspm.fr (GS);  
jean@astro.ufsc.br (JMG)

information from observational data in a reliable way is crucial for a deeper understanding of galaxy formation and evolution.

The mapping of observed on to physical properties of galaxies has been a major topic of research for over a generation of astronomers since the pioneering works of Morgan (1956), Wood (1966) and Faber (1972), on the one hand, and Tinsley (1968) and Spinrad & Taylor (1972), on the other. The first group of authors introduced the so-called empirical population synthesis methods, which aim to reproduce a set of observations of a given galaxy by means of a linear combination of simpler systems of known characteristics, like individual stars or chemically homogeneous and coeval groups of stars. Bica (1988), Pelat (1997), Cid Fernandes et al. (2001) and Moulataka et al. (2004) are examples of modern studies following this approach. The second group pioneered the so-called evolutionary population synthesis methods, which compare galaxy data with models that follow the time evolution of an entire stellar system by combining libraries of evolutionary tracks and stellar spectra with prescriptions for the initial mass function (IMF), star formation and chemical histories. This approach has enjoyed more widespread use in recent years, e.g. Arimoto & Yoshii (1987), Guiderdoni & Rocca-Volmerange (1987), Bressan, Chiosi & Tantalo (1996), Fioç & Rocca-Volmerange (1997), Vazdekis (1999), Bruzual & Charlot (2003, hereafter BC03) and Le Borgne et al. (2004), among many others (see Cardiel et al. 2003 for a large set of references). In short, empirical synthesis relies on nature for its basic ingredients, whereas evolutionary synthesis relies mostly on models. However, since models are made and calibrated to mimic nature, this difference is gradually vanishing as models improve.

Besides different methodologies, there are also differences in what type of data is actually modelled. Colours (e.g. Wood 1966), absorption-line equivalent widths or spectral indices (e.g. Worthey 1994; Kauffmann et al. 2003, hereafter K03) and emission features, both stellar (Leitherer, Robert & Heckman 1995; Schaerer & Vacca 1998) and nebular (Mas-Hesse & Kunth 1991; Kewley et al. 2001), have all been used in stellar population synthesis. More recently, the full spectral information has been incorporated in the modelling process, both including (Charlot & Longhetti 2001) and excluding (Vazdekis & Arimoto 1999; Reichardt, Jimenez & Heavens 2001; Cid Fernandes et al. 2004, hereafter CF04) emission lines.

Recovering the stellar content of a galaxy from its observed integrated spectrum is not an easy task, as can be deduced from the amount of work devoted to this topic over the past half century or so. The situation is, however, much more favourable nowadays. Huge observational and theoretical efforts in the past few years have produced large sets of high-quality spectra of stars (e.g. Prugniel & Soubiran 2001; Le Borgne et al. 2003; Bertone et al. 2004; González et al. 2005). These libraries are being implemented in a new generation of evolutionary synthesis models, allowing the prediction of galaxy spectra with an unprecedented level of detail (Vazdekis 1999; BC03; Le Borgne et al. 2004). At the same time, galaxy spectra are now more abundant than ever (Loveday et al. 1996; York et al. 2000). The Sloan Digital Sky Survey (SDSS), in particular, is providing a homogeneous data base of *hundreds of thousands* of galaxy spectra in the 3800–9200 Å range, with a resolution of  $\lambda/\Delta\lambda \sim 1800$  (York et al. 2000; Stoughton et al. 2002; Abazajian et al. 2003, 2004). This enormous amount of high-quality data will undoubtedly be at the heart of tremendous progress in our understanding of galaxy constitution, formation and evolution. Indeed, significant steps in this direction have recently been made (K03; Brinchmann et al. 2004; Tremonti et al. 2004; Heavens et al. 2004; Panter, Heavens & Jimenez 2004).

In order to take advantage of the recent progress in evolutionary synthesis to analyse data sets such as the SDSS, a methodology must be set up to go from the observed spectra to physical properties of galaxies. In this paper we discuss one possible method to achieve this goal. The method is based on fitting an observed spectrum with a linear combination of simple theoretical stellar populations (coeval and chemically homogeneous) computed with evolutionary synthesis models at the same spectral resolution as that of the SDSS.

Our goal here is to demonstrate that, besides providing excellent starlight templates to aid emission-line studies, spectral synthesis recovers reliable stellar population properties out of galaxy spectra of realistic quality. We show that this simple method provides robust information on the distributions of stellar age ( $t_*$ ) and stellar metallicity ( $Z_*$ ), as well as on the extinction, velocity dispersion and stellar mass. The ability to recover information on  $Z_*$  is particularly welcome, given that stellar metallicities are notoriously more difficult to assess than other properties. In order to reach this goal, we follow: (1) a priori arguments, based on simulations; (2) comparisons with independent work based on a different method; and (3) an a posteriori empirical analysis of the consistency of results obtained for a large sample of SDSS galaxies. Other papers in this series on the semi-empirical analysis of galaxies (SEAGal) will explore various astrophysical implications of the results obtained with this method.

This paper is organized as follows. Section 2 presents an overview of our synthesis method and simulations designed to test it and evaluate the uncertainties involved. The discussion is focused on how to use the synthesis to derive robust estimators of physically interesting stellar population properties. Section 3 defines a volume-limited sample of SDSS galaxies and presents the results of the synthesis of their spectra, along with measurements of emission lines. In Section 4 we compare our results to those recently published by Brinchmann et al. (2004). Stellar population and emission-line properties are used in Section 5 to investigate whether the synthesis produces astrophysically plausible results. Finally, Section 6 summarizes our main findings.

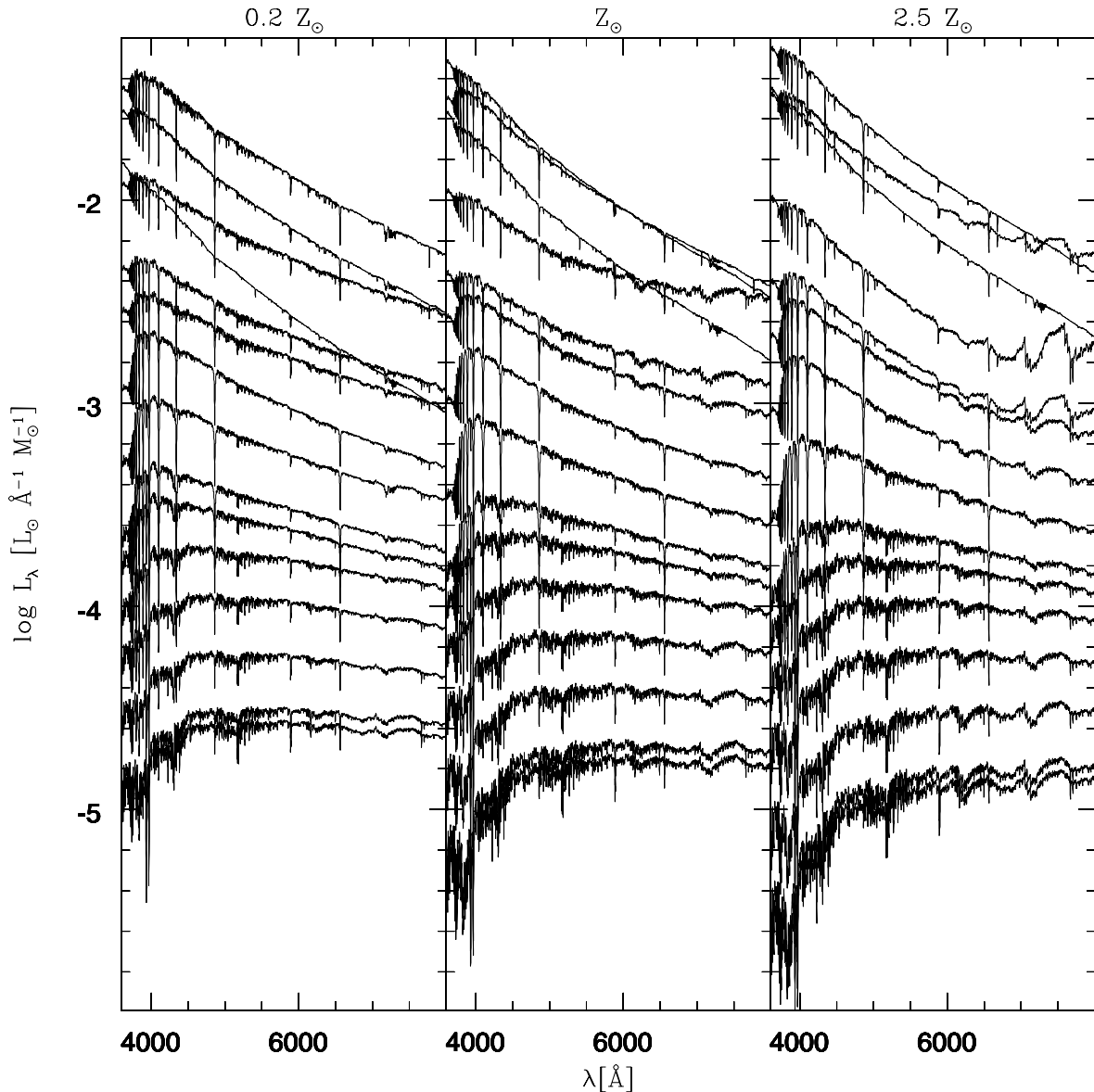
## 2 SPECTRAL SYNTHESIS

### 2.1 Method

Our synthesis code, which we call STARLIGHT, was first discussed in CF04 [see also Cid Fernandes et al. (2005) and Garcia-Rissmann et al. (2005) for different applications of the same code]. STARLIGHT mixes computational techniques originally developed for empirical population synthesis with ingredients from evolutionary synthesis models. Briefly, we fit an observed spectrum  $O_\lambda$  with a combination of  $N_*$  simple stellar populations (SSP) from the evolutionary synthesis models of BC03. Extinction is modelled as due to foreground dust, and parametrized by the  $V$ -band extinction  $A_V$ . The Galactic extinction law of Cardelli, Clayton & Mathis (1989) with  $R_V = 3.1$  is adopted. Line-of-sight stellar motions are modelled by a Gaussian distribution  $G$  centred at velocity  $v_*$  and with dispersion  $\sigma_*$ . With these assumptions, the model spectrum is given by

$$M_\lambda = M_{\lambda_0} \left( \sum_{j=1}^{N_*} x_j b_{j,\lambda} r_\lambda \right) \otimes G(v_*, \sigma_*), \quad (1)$$

where  $b_{j,\lambda}$  is the spectrum of the  $j$ th SSP normalized at  $\lambda_0$ ,  $r_\lambda \equiv 10^{-0.4(A_\lambda - A_{\lambda_0})}$  is the reddening term,  $M_{\lambda_0}$  is the synthetic flux at the normalization wavelength,  $\mathbf{x}$  is the *population vector* and  $\otimes$  denotes the convolution operator. Each component  $x_j$  ( $j = 1, \dots, N_*$ )



**Figure 1.** Spectra of the 45 SSPs used in the spectral synthesis (from BC03). The base comprises three different metallicities,  $Z = 0.2, 1$  and  $2.5 Z_{\odot}$ , and 15 ages: from top to bottom,  $t = 0.001, 0.00316, 0.00501, 0.01, 0.02512, 0.04, 0.10152, 0.28612, 0.64054, 0.90479, 1.434, 2.5, 5, 11$  and  $13$  Gyr. All SSPs are normalized to  $1 M_{\odot}$  at  $t = 0$ .

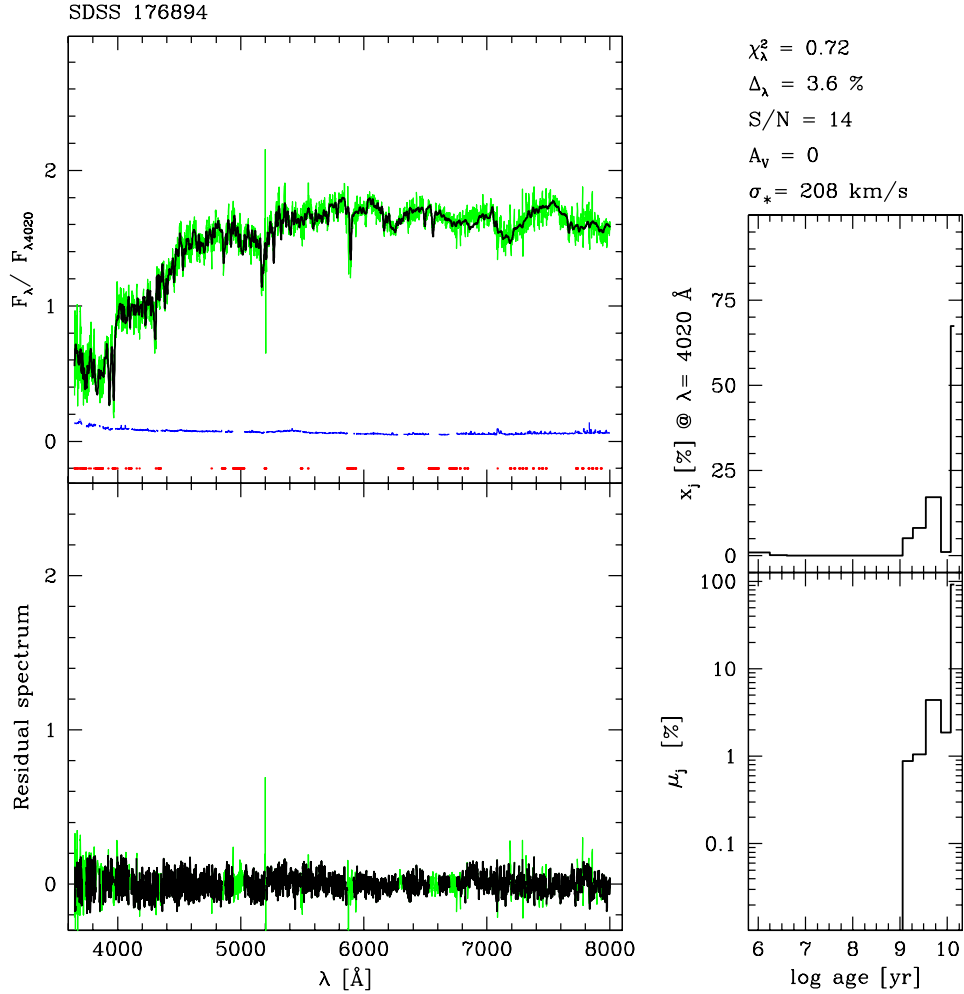
represents the fractional contribution of the SSP with age  $t_j$  and metallicity<sup>1</sup>  $Z_j$  to the model flux at  $\lambda_0$ . The base components can be equivalently expressed as a mass fraction vector  $\mu$ . In this work we adopt a base with  $N_* = 45$  SSPs, encompassing 15 ages between  $10^6$  and  $1.3 \times 10^{10}$  yr and three metallicities,  $Z = 0.2, 1$  and  $2.5 Z_{\odot}$ . Their spectra, shown in Fig. 1, were computed with the STELIB library (Le Borgne et al. 2003), Padova-2004 models, and Chabrier (2003) IMF (see BC03 for details).

The fit is carried out with a simulated annealing plus Metropolis scheme, which searches for the minimum  $\chi^2 = \sum_{\lambda} [(O_{\lambda} - M_{\lambda}) w_{\lambda}]^2$ , where  $w_{\lambda}^{-1}$  is the error in  $O_{\lambda}$ . Regions around emission lines, bad pixels or sky residuals are masked out by setting  $w_{\lambda} = 0$ . Pixels that

deviate by more than three times the rms between  $O_{\lambda}$  and an initial estimate of  $M_{\lambda}$  are also given zero weight.

The minimization consists of a series of  $N_M = 6$  likelihood-guided Metropolis explorations of the parameter space. From each iteration to the next we increase the  $w_{\lambda}$  weights geometrically [which corresponds to a decrease in the ‘temperature’ in statistical mechanics terms (e.g. MacKay 2003)]. The step size in each parameter is concomitantly decreased and the number of steps is increased. This scheme gradually focuses on the most likely region in parameter space, avoiding (through the logic of the Metropolis algorithm) trapping on to local minima. After completion, the whole fit is finetuned repeating the full loop excluding all  $x_j = 0$  components. An important difference (from the computational point of view) with respect to the code in CF04 is that we now perform a series expansion of the  $r_{\lambda}$  extinction factor, which allows a much faster computation of  $\chi^2$ . Naturally, there are a number of technical parameters in this complex fitting algorithm, which we have optimized by means of

<sup>1</sup> In this paper we follow the convention used in stellar evolution studies, which define stellar metallicities in terms of the fraction of mass in metals. In this system, the Sun has  $Z_* = 0.02$ .



**Figure 2.** Spectral synthesis of an early-type SDSS galaxy. Top left: Observed (thin line, green online), model (thick line, black online) and error spectra (dashed line, blue online). Points at the bottom indicate bad pixels (as given by the SDSS flag) or emission-line windows, both of which were masked out in the fits. Bottom left: Residual spectrum. Masked regions are plotted with a thin line (green online). Right: Flux (top) and mass (bottom) fractions as a function of age. Some of the derived properties are listed in the top right:  $\chi_\lambda^2$  is the reduced  $\chi^2$ ;  $\Delta_\lambda$  is the mean relative difference between model and observed spectra; S/N refers to the region around  $\lambda_0 = 4020 \text{ \AA}$ .

extensive simulations (see CF04). At any rate, results reported here are robust with respect to variations in these technical parameters.

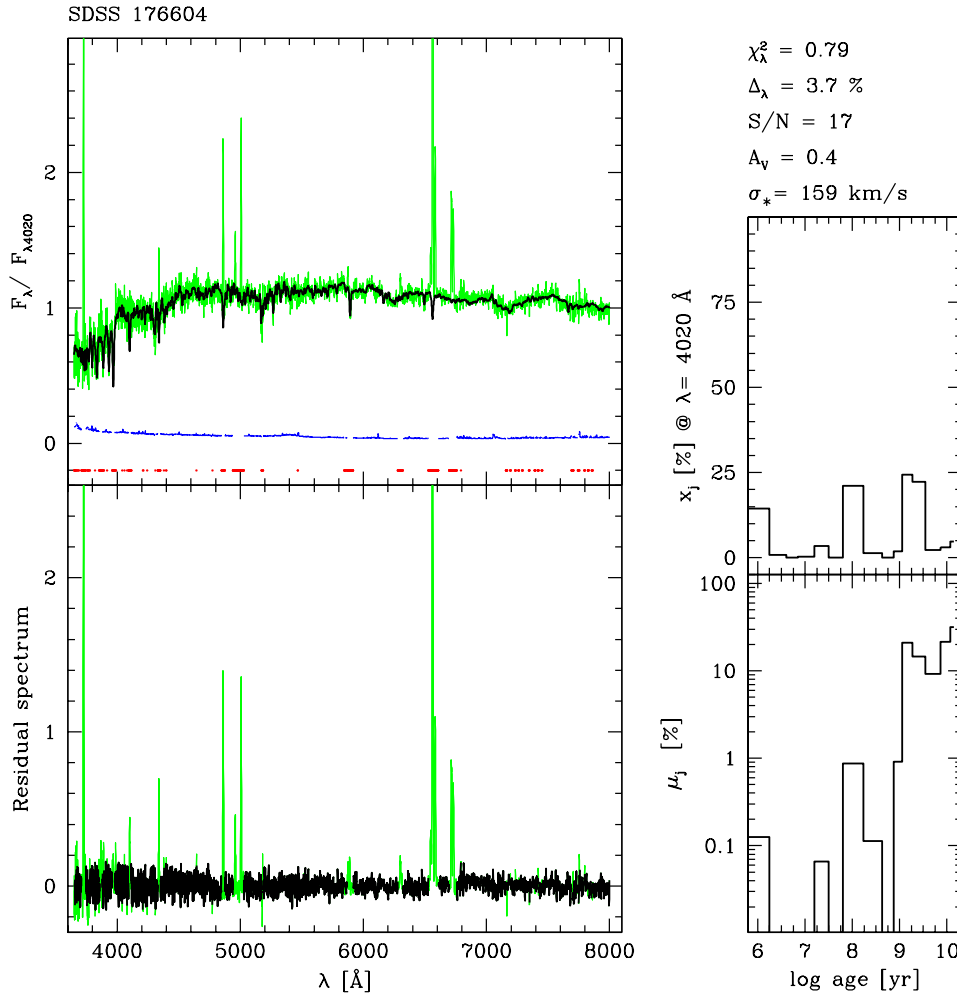
Figs 2 and 3 illustrate the spectral fits obtained for two galaxies drawn from the SDSS data base. The top-left panel shows the observed spectrum (thin line) and the model (thick), as well as the error spectrum (dashed). The bottom-left panel shows the  $O_\lambda - M_\lambda$  residual spectrum, while the panels on the right summarize the derived star formation history encoded in the age-binned population vector. These examples, along with those in K03 and CF04, demonstrate that this simple method is capable of reproducing real galaxy spectra to an excellent degree of accuracy.

An important application of the synthesis is to measure emission lines from the residual spectrum, as done by K03 (see also Section 3.3). Another, of course, is to infer stellar population properties from the fitting parameters. This was not the approach followed by K03 and their subsequent papers, who prefer to derive stellar population properties mainly from the  $D_n(4000)$  versus  $H\delta_\Lambda$  diagram. Our central goal here is to investigate whether spectral synthesis can also recover reliable stellar population properties. In the remainder of this section, we address this issue by means of simulations.

## 2.2 Robust description of the synthesis results

The existence of multiple solutions is an old known problem in stellar population synthesis. This multiplicity arises from a combination of three factors: (1) algebraic degeneracy (number of unknowns larger than number of observables), (2) intrinsic degeneracies of stellar populations, and (3) measurement uncertainties. Unlike in methods that synthesize only a handful of spectral indices, algebraic degeneracy is not a problem for the method outlined above, as the number of  $O_\lambda$  points in any decently sampled spectrum far exceeds the number of parameters even for large bases. Similarly, by modelling the whole spectrum, one should be able to alleviate degeneracies in spectral indices associated with different stellar populations (Jimenez et al. 2004). Measurement errors, however, are still a problem. A corollary of these complications is that even superb spectral fits such as those shown in Figs 2 and 3 do not guarantee that the resulting parameter estimates are trustworthy.

This brief discussion illustrates the need to assess the degree to which one can trust the parameters involved in the fit before using them to infer stellar population properties. As posed above, the



**Figure 3.** As Fig. 2, but for a late-type galaxy.

spectral fits involve  $N_* + 3$  parameters:  $N_* - 1$  of the  $\mathbf{x}$  components (one degree of freedom is removed by the normalization constraint),  $M_{\lambda_0}$ ,  $A_V$  and the two kinematical parameters,  $v_*$  and  $\sigma_*$ . The reliability of parameter estimation is best studied by means of simulations that feed the code with spectra generated with known parameters, add noise, and then examine the correspondence between input and output values.

CF04 performed this kind of simulation for an  $N_* = 20$  base and spectra in the 3500–5200 Å interval. Their analysis concentrated on how well the method recovers  $\mathbf{x}_{\text{input}}$  for varying degrees of noise. The main results of that study are as follows: (1) In the absence of noise, the method recovers all components of  $\mathbf{x}$  to a high degree of accuracy. (2) In the presence of noise, however, the individual output  $x_j$  fractions may deviate drastically from the input values. Essentially, what happens is that noise washes away the differences between spectrally similar components, so it becomes impossible to distinguish them, and the code splits  $x_j$  among neighbouring components in spectral space. This is a common problem in population synthesis (e.g. Cid Fernandes et al. 2001; Tremonti 2003; Panter et al. 2004).

Once we have identified the origin of the problem, the remedy to fix it is evident: bin over spectrally similar components. In other words, instead of attempting a fine-grained description of stellar population mixtures in terms of many ages and metallicities, it is

much better to ‘marginalize over the details’ and work with a coarser but more robust description based on combined  $x_j$  fractions. As shown by CF04, condensed versions of the population vector that project its  $N_*$  components on to just a few physically interesting axes yield very reliable results.

Our approach and objectives are conceptually similar to those of the MOPED code of Heavens et al. (2000, see also Reichardt et al. 2001; Panter et al. 2004). Operationally, however, the two codes differ. MOPED compresses data on input by degrading the SDSS spectral resolution to 20 Å and performing weighted linear combinations of  $O_\lambda$ , reducing its dimension to one datum per output parameter while at the same time minimizing the loss of information. STARLIGHT, on the other hand, works with the spectrum at its full resolution and an overdimensioned set of parameters (i.e. large  $N_*$ ), which we compress on the description of the output. The disadvantage of our approach is computational. Indeed, it is described as a ‘slow’ and ‘brute force’ method by Panter et al. (2004). This, however, is not a severe problem given the abundance of CPUs nowadays (STARLIGHT takes about 4 min per galaxy on a 2 GHz Linux workstation). Furthermore, there is plenty of room for improvement in the efficiency of the algorithm – for instance, by implementing a smarter exploration of the parameter space (Doran & Müller 2004), to the point that computational constraints could soon become a minor concern. An advantage of STARLIGHT is that it also measures kinematical

parameters (mainly  $\sigma_*$ ) and provides a high-resolution template spectrum.

### 2.2.1 Simulations

We have carried out new simulations designed to test the method and investigate which combinations of the parameters provide robust results. These simulations differ from those in CF04 in three main aspects: (1) the spectral range is now 3650–8000 Å; (2) the new base is larger; and (3) a more realistic error spectrum was used. A further difference is that we use a larger number of iterations, partly because we have more base elements than CF04 and partly because the code is now over 200 times faster. At each step in the annealing scheme, the number of steps per parameter is set to  $2(L/\epsilon)^2$ , where  $L$  is the length of the allowed range for the parameter (e.g.  $L = 1$  for  $x_j$  fractions, which range from 0 to 1), and  $\epsilon$  is the step size (see CF04 for a more detailed description of these technical aspects). This ensures that each parameter can in principle random-walk twice its whole allowed range, which is an adequate criterion for Metropolis sampling (e.g. MacKay 2003). Furthermore, there are 12 stages in the annealing schedule. Overall, over  $10^7$  combinations of parameters are sampled for each galaxy.

Several sets of simulations were performed. Given our interest in modelling SDSS galaxies, here we focus on simulations tailored to match the characteristics of this data set. Test galaxies were built from the average  $\mathbf{x}$ ,  $A_V$  and  $\sigma_*$  within 65 boxes in the mean stellar age versus mean stellar metallicity plane obtained for the sample described in Section 3. Each spectrum covers the 3650–8000 Å range in steps of 1 Å, the same sampling as the BC03 models. We generate 20 perturbed versions of each synthetic spectrum for each of five levels of noise: signal-to-noise ratio  $S/N = 5, 10, 15, 20$  and

30, where  $S/N$  is the signal-to-noise ratio  $\text{Å}^{-1}$  in the region around  $\lambda_0 = 4020 \text{ Å}$ . Unlike in CF04, who used a flat error spectrum, the error at each  $\lambda$  was assumed to follow a Gaussian distribution with amplitude obtained by scaling a normalized mean SDSS error spectrum to yield the desired  $S/N$  at  $\lambda_0$ . This error spectrum decreases by a factor of  $\sim 3$  from 3650 to 6200 Å and then increases again by  $\sim 1.5$  towards 8000 Å. Finally, as for the actual data fits, we mask points around [O II]  $\lambda\lambda 3726, 3729$ , [Ne III]  $\lambda 3869$ , He, H $\delta$ , H $\gamma$ , H $\beta$ , [O III]  $\lambda\lambda 4959, 5007$ , [He I]  $\lambda 5876$ , NaD  $\lambda 5890$ , [O I]  $\lambda 6300$ , [N II]  $\lambda\lambda 6548, 6583$ , H $\alpha$ , and [S II]  $\lambda\lambda 6717, 6731$ .

These new simulations confirm that the individual components of  $\mathbf{x}$  are very uncertain, so we skip a detailed comparison between  $\mathbf{x}_{\text{input}}$  and  $\mathbf{x}_{\text{output}}$  and jump straight to results based on more robust descriptions of the synthesis output.

### 2.2.2 Condensed population vector

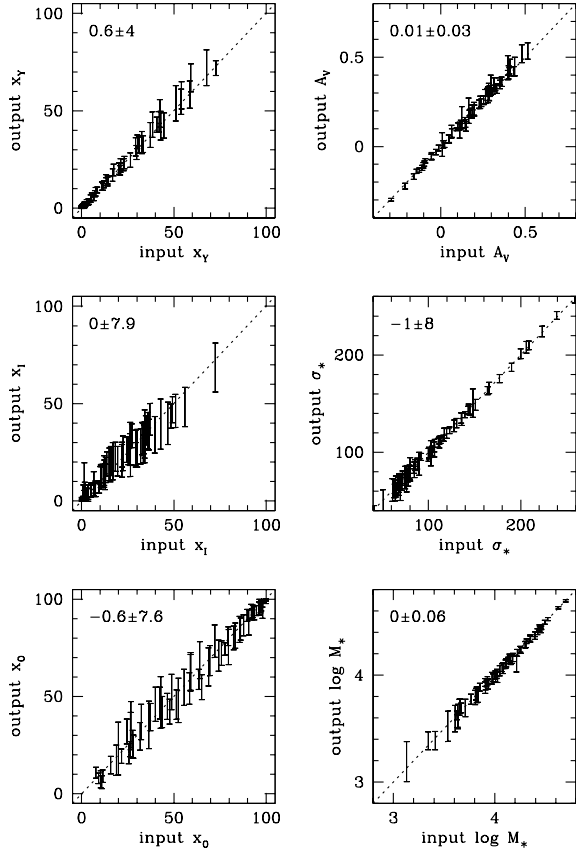
A coarse but robust description of the star formation history of a galaxy may be obtained by binning  $\mathbf{x}$  on to ‘young’ ( $t_j < 10^8 \text{ yr}$ ), ‘intermediate-age’ ( $10^8 \leq t_j \leq 10^9 \text{ yr}$ ), and ‘old’ ( $t_j > 10^9 \text{ yr}$ ) components ( $x_Y, x_I$  and  $x_O$ , respectively). These age ranges were defined on the basis of the simulations, by seeking which combinations of the  $x_j$  produce smaller input minus output residuals. Table 1 and Fig. 4 show that these three components are very well recovered by the method, with uncertainties smaller than  $\Delta x_Y = 0.05$ ,  $\Delta x_I = 0.1$  and  $\Delta x_O = 0.1$  for  $S/N \geq 10$ .

### 2.2.3 Mass, extinction and velocity dispersion

Fig. 4 also shows the input versus output values of  $A_V$ ,  $\sigma_*$  and the stellar mass  $M_*$ . The latter is not an explicit input parameter of

**Table 1.** Summary of parameter uncertainties. Each row corresponds to a parameter constructed by combining the parameters given by the synthesis. The different columns list the mean  $\pm$  rms difference between output and input values of the corresponding quantity, as obtained from simulations with different signal-to-noise ratios. For instance, the mean output minus input difference in  $(\log t_*)_L$  for  $S/N = 10$  is 0.01 dex and the rms dispersion is 0.08 dex. The last three rows describe the statistics of goodness-of-fit indicators. The units are percentage for light and mass fractions ( $x_Y, \dots, \mu_O$ ), mag for  $A_V$ ,  $\text{km s}^{-1}$  for  $v_*$  and  $\sigma_*$ , percentage for  $\Delta_\lambda$ , and dex for logarithmic quantities.

Parameter	S/N at $\lambda = 4020 \text{ Å}$				
	5	10	15	20	30
$x_Y$	0.35 $\pm$ 7.06	0.21 $\pm$ 4.50	0.62 $\pm$ 4.04	0.56 $\pm$ 3.04	0.64 $\pm$ 2.63
$x_I$	0.76 $\pm$ 13.94	−0.06 $\pm$ 9.00	0.01 $\pm$ 7.88	−0.22 $\pm$ 6.26	−0.02 $\pm$ 5.07
$x_O$	−1.12 $\pm$ 13.02	−0.15 $\pm$ 8.83	−0.63 $\pm$ 7.61	−0.34 $\pm$ 6.19	−0.62 $\pm$ 5.05
$\mu_Y$	0.02 $\pm$ 1.66	0.11 $\pm$ 1.41	0.16 $\pm$ 1.18	0.18 $\pm$ 1.05	0.19 $\pm$ 0.80
$\mu_I$	1.11 $\pm$ 10.26	0.99 $\pm$ 7.57	0.93 $\pm$ 6.10	1.00 $\pm$ 5.03	1.17 $\pm$ 4.22
$\mu_O$	−1.13 $\pm$ 10.54	−1.10 $\pm$ 8.17	−1.09 $\pm$ 6.54	−1.18 $\pm$ 5.59	−1.36 $\pm$ 4.63
$A_V$	0.01 $\pm$ 0.09	0.00 $\pm$ 0.05	0.01 $\pm$ 0.03	0.01 $\pm$ 0.03	0.00 $\pm$ 0.02
$v_*$	0.24 $\pm$ 17.73	−0.11 $\pm$ 8.55	0.07 $\pm$ 5.92	0.00 $\pm$ 4.23	0.05 $\pm$ 2.81
$\sigma_*$	−3.12 $\pm$ 24.32	−1.40 $\pm$ 12.36	−1.01 $\pm$ 7.71	−0.75 $\pm$ 5.73	−0.57 $\pm$ 3.78
$\log M_*$	0.01 $\pm$ 0.11	−0.01 $\pm$ 0.08	−0.01 $\pm$ 0.06	−0.01 $\pm$ 0.05	−0.02 $\pm$ 0.04
$\langle \log t_* \rangle_L$	−0.01 $\pm$ 0.14	0.01 $\pm$ 0.08	0.00 $\pm$ 0.06	0.01 $\pm$ 0.05	0.01 $\pm$ 0.04
$\langle \log t_* \rangle_M$	−0.03 $\pm$ 0.20	−0.04 $\pm$ 0.14	−0.03 $\pm$ 0.11	−0.04 $\pm$ 0.10	−0.04 $\pm$ 0.08
$\log(Z_*)_L$	−0.01 $\pm$ 0.15	−0.01 $\pm$ 0.09	0.00 $\pm$ 0.08	0.00 $\pm$ 0.06	0.00 $\pm$ 0.05
$\log(Z_*)_M$	−0.03 $\pm$ 0.18	−0.02 $\pm$ 0.13	−0.03 $\pm$ 0.11	−0.02 $\pm$ 0.09	−0.03 $\pm$ 0.08
$\sigma_L(\log t_*)$	0.03 $\pm$ 0.16	0.00 $\pm$ 0.10	0.00 $\pm$ 0.08	−0.01 $\pm$ 0.07	−0.01 $\pm$ 0.06
$\sigma_M(\log t_*)$	−0.04 $\pm$ 0.09	−0.02 $\pm$ 0.06	−0.01 $\pm$ 0.05	−0.01 $\pm$ 0.04	0.00 $\pm$ 0.04
$\sigma_L(Z_*)$	−0.01 $\pm$ 0.21	0.00 $\pm$ 0.14	0.02 $\pm$ 0.11	0.02 $\pm$ 0.10	0.02 $\pm$ 0.09
$\sigma_M(Z_*)$	−0.11 $\pm$ 0.26	−0.05 $\pm$ 0.19	−0.03 $\pm$ 0.16	−0.02 $\pm$ 0.14	−0.02 $\pm$ 0.12
$\chi^2/N_\lambda$	0.95 $\pm$ 0.02	0.95 $\pm$ 0.02	0.95 $\pm$ 0.02	0.95 $\pm$ 0.02	0.95 $\pm$ 0.02
$\Delta_\lambda$	11.64 $\pm$ 4.05	5.37 $\pm$ 1.19	3.54 $\pm$ 0.78	2.65 $\pm$ 0.58	1.76 $\pm$ 0.39
$\Delta_\lambda \times (S/N)_{\lambda_0}$	0.78 $\pm$ 0.27	0.68 $\pm$ 0.20	0.64 $\pm$ 0.21	0.59 $\pm$ 0.22	0.51 $\pm$ 0.23



**Figure 4.** Input versus output synthesis parameters for simulations with  $S/N = 15$  at  $\lambda_0 = 4020 \text{ \AA}$ . The units are percentages for the condensed population vector ( $x_Y$ ,  $x_I$ ,  $x_O$ ), magnitudes for  $A_V$ ,  $\text{km s}^{-1}$  for  $\sigma_*$  and arbitrary for  $M_*$ . The output parameters are represented by  $\pm 1\sigma$  error bars centred on the mean values obtained by fitting 20 realizations of each of 65 test galaxies.

the models, but may be computed from  $\mu$  and the  $M_*/L_{\lambda_0}$  ratio of the different populations in the base. The uncertainties in these parameters are  $\Delta A_V < 0.05 \text{ mag}$ ,  $\Delta \log M_* < 0.1 \text{ dex}$  and  $\Delta \sigma_* < 12 \text{ km s}^{-1}$  for  $S/N \geq 10$  (Table 1).

#### 2.2.4 Mean stellar age

If one had to choose a single parameter to characterize the stellar population mixture of a galaxy, the option would certainly be its mean age. We define two versions of mean stellar age (the logarithm of the age, actually), one weighted by light,

$$\langle \log t_* \rangle_L = \sum_{j=1}^{N_*} x_j \log t_j, \quad (2)$$

and another weighted by stellar mass,

$$\langle \log t_* \rangle_M = \sum_{j=1}^{N_*} \mu_j \log t_j. \quad (3)$$

Note that, by construction, both definitions are limited to the 1 Myr–13 Gyr range spanned by the base. The mass-weighted mean age is in principle more physical, but, because of the non-constant  $M/L$  of stars, it has a much less direct relation with the observed spectrum than  $\langle \log t_* \rangle_L$ .

Fig. 5 shows the input against output  $\langle \log t_* \rangle_L$  for simulations with  $S/N = 10$  and 20. The plots show that the mean age is a

very robust quantity. The rms difference between input and output  $\langle \log t_* \rangle_L$  values is  $\leq 0.08 \text{ dex}$  for  $S/N > 10$ , and  $\leq 0.14 \text{ dex}$  for  $\langle \log t_* \rangle_M$  (Table 1). Although the uncertainties of  $\langle \log t_* \rangle_L$  and  $\langle \log t_* \rangle_M$  are comparable in absolute terms, the latter index spans a smaller dynamical range (because of the large  $M/L$  ratio of old populations), so in practice  $\langle \log t_* \rangle_L$  is the more useful of the two indices.

Given that the mean stellar age is so well recovered by the method, one might attempt more detailed descriptions of the star formation history involving, say, higher moments of the age distribution. For instance,

$$\sigma_L(\log t_*) = \left[ \sum_{j=1}^{N_*} x_j (\log t_j - \langle \log t_* \rangle_L)^2 \right]^{1/2} \quad (4)$$

measures the flux-weighted standard deviation of the log age distribution, and might be useful to distinguish galaxies dominated by a single population from those which had continuous or bursty star formation histories. The uncertainty in this index is of the order of 0.1 dex.

#### 2.2.5 Mean stellar metallicity

Given an option of what to choose as a second parameter to describe a mixed stellar population, the choice would likely be its typical metallicity. Analogously to what we did for ages, we define both light- and mass-weighted mean stellar metallicities:

$$\langle Z_* \rangle_L = \sum_{j=1}^{N_*} x_j Z_j \quad (5)$$

and

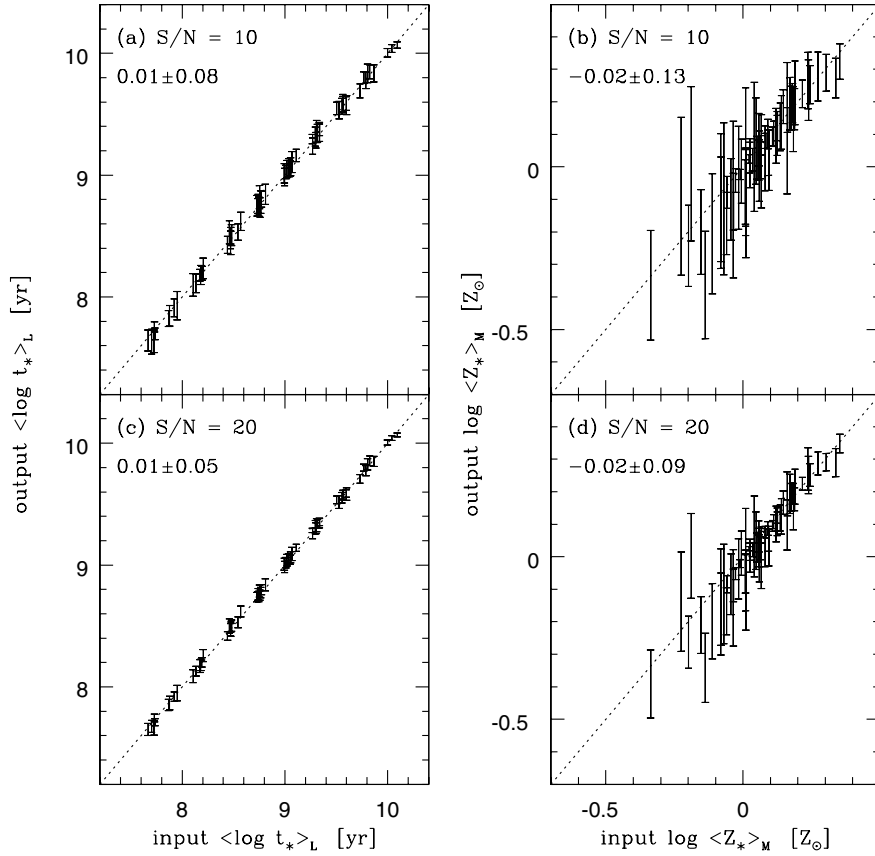
$$\langle Z_* \rangle_M = \sum_{j=1}^{N_*} \mu_j Z_j, \quad (6)$$

both of which are bounded by the 0.2–2.5  $Z_\odot$  base limits. Fig. 5 and Table 1 show that the rms of  $\Delta \log \langle Z_* \rangle_M = \log \langle Z_* \rangle_{M,\text{output}} - \log \langle Z_* \rangle_{M,\text{input}}$  is of the order of 0.1 dex. In absolute terms this is comparable to  $\Delta \langle \log t_* \rangle$ , but note that  $\langle Z_* \rangle$  covers a much narrower dynamical range than  $\langle \log t_* \rangle$ , so that in practice mean stellar metallicities are more sensitive to errors than are mean ages. This is not surprising, given that age is the main driver of variance among SSP spectra, metallicity having a ‘second-order’ effect (e.g. Schmidt et al. 1991; Ronen, Aragon-Salamanca & Lahav 1999). This is the reason why studies of the stellar populations of galaxies have a much harder time estimating metallicities than ages, to the point that one is often forced to bin over the  $Z$  information and deal only with age-related estimates such as  $\langle \log t_* \rangle$  (e.g. Cid Fernandes et al. 2001; Cid Fernandes, Leão & Lacerda 2003; K03).

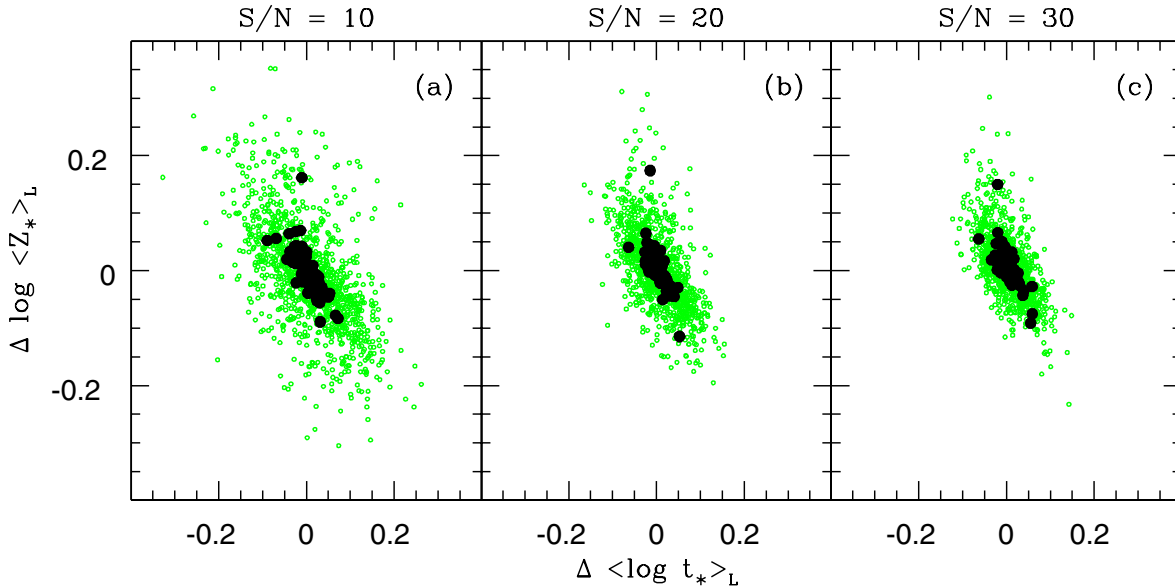
Notwithstanding these notes, it is clear that uncertainties of  $\sim 0.1 \text{ dex}$  in  $\langle Z_* \rangle$  are actually good news, since they do allow us to recover useful information on an important but hard-to-measure property. This new tracer of stellar metallicity is best applicable to large samples of galaxies such as the SDSS. The statistics of samples help in reducing uncertainties associated with  $\langle Z_* \rangle$  estimates for single objects and allows one to investigate correlations between  $\langle Z_* \rangle$  and other galaxy properties (Sodr e et al., in preparation; Section 5.1).

#### 2.2.6 Age–metallicity degeneracy

Our method tends to underestimate  $\langle Z_* \rangle$  for metal-rich systems and vice versa. This bias is due to the infamous age–metallicity degeneracy. In order to verify the degree to which our synthesis



**Figure 5.** Input versus output mean stellar ages and metallicities for simulations with S/N = 10 and 20.



**Figure 6.** Output minus input residuals in  $\log \langle Z_* \rangle_L$  against residuals in  $\langle \log t_* \rangle_L$ . Small circles (green online) are the individual test galaxies. Larger filled black circles mark the mean residuals for 20 perturbed versions of the same test galaxy. The anticorrelation illustrates the effect of the age–metallicity degeneracy.

is affected by this well-known problem (e.g. Renzini & Buzzoni 1986; Worthey 1994; Bressan, Chiosi & Tantalo 1996), we have examined the correlation between the output minus input residuals in  $\langle \log t_* \rangle_L$  and  $\log \langle Z_* \rangle_L$ . The age– $Z$  degeneracy acts in the sense of confusing old, metal-poor systems with young, metal-rich

ones and vice versa, which should produce anticorrelated residuals. These are indeed seen in Fig. 6, where small symbols represent all 1300 individual simulations and larger filled circles correspond to the mean residuals obtained for each set of 20 perturbations of the 65 test galaxies. This anticorrelation is also present with the



mass-weighted mean stellar metallicity  $\langle Z_{\star} \rangle_M$ , but is not as strong as for  $\langle Z_{\star} \rangle_L$ . On the other hand, the uncertainty in  $\langle Z_{\star} \rangle_M$  is always larger than for  $\langle Z_{\star} \rangle_L$  (Table 1).

The age– $Z$  degeneracy is thus present in our method, introducing systematic biases in our  $\langle Z_{\star} \rangle$  and  $\langle \log t_{\star} \rangle$  estimates at the level of up to  $\sim 0.1$ – $0.2$  dex. None of the results reported in this paper rely on this level of precision.

### 2.3 Simulations for random parameters

We have also carried out a separate set of simulations for 100 test galaxies generated by random combinations of  $x$ ,  $A_V$  and  $\sigma_{\star}$ . These simulations differ from those presented above in that the test galaxies are not restricted to match the range of properties inferred from the application of the synthesis to SDSS galaxies (Section 3). Given that these random galaxies span a broader (but less representative of our sample galaxies) region of the parameter space, one might expect to find larger uncertainties in the derived properties. This was indeed confirmed in the numerical experiments, although the effect is small. The uncertainties in the  $(x_Y, x_I, x_O)$  binned fractions increase by just a couple of percentage points with respect to those in Table 1. Similarly, the increase in  $\Delta A_V$  is of just 0.02 mag, while  $\Delta \langle \log t_{\star} \rangle_L$  and  $\Delta \langle \log Z_{\star} \rangle_M$  increase by  $\sim 0.02$  and 0.03 dex, respectively.

The only properties whose uncertainties are substantially larger than those reported in Table 1 are the stellar mass and velocity dispersion. For instance, while the SDSS-based simulations for  $S/N = 10$  yield  $\Delta \log M_{\star} = 0.08$  dex and  $\Delta \sigma_{\star} = 12$  km s $^{-1}$ , with random galaxies these errors increase to 0.15 dex and 24 km s $^{-1}$ , respectively. The reason for this apparent discrepancy is due to the fact that the set of random test galaxies contains a larger proportion of systems dominated by very young stellar populations than we find for SDSS galaxies. An error  $\Delta x_Y$  in the light fraction associated to these populations must be compensated by errors in the older components, which carry most of the mass, even when  $x_Y$  is large. As illustrated by the size of the error bars in Fig. 4,  $\Delta x_Y$  increases as  $x_Y$  increases, so the errors in the mass fraction components increase too, leading to larger dispersion in  $M_{\star}$ . Furthermore, galaxies with large  $x_Y$  have few absorption features to constrain the kinematical broadening of the spectrum, which explains the larger dispersion in our  $\sigma_{\star}$  estimates. To prove this point, we have re-evaluated the uncertainties in  $M_{\star}$  and  $\sigma_{\star}$  excluding test galaxies with  $x_Y > 70$  per cent, which corresponds to systems that formed  $\gtrsim 20$  per cent of their stellar mass over the past  $< 10^8$  yr. For this subset of the simulations, which comprise 72 galaxies (each one split on to 20 different spectra corresponding to independent Monte Carlo realizations of the noise), the uncertainties in mass and velocity dispersion decrease to  $\Delta \log M_{\star} = 0.09$  dex and  $\Delta \sigma_{\star} = 14$  km s $^{-1}$  (for  $S/N = 10$ ), only slightly larger than those reported in Section 2.2.3. Uncertainties in other properties also decrease to values very similar to those listed in Table 1.

We thus conclude that the parameter uncertainties studied in Section 2.2 and summarized in Table 1 are only moderately affected by the design of the simulations, and represent fair estimates of the limitations of our synthesis method.

Finally, we have carried out simulations using the  $Z = 0.02 Z_{\odot}$  BC03 SSPs to generate further test galaxies. Galaxies with such low metallicity are not expected to be present in significant numbers in the sample described in Section 3.1, given that it excludes low-luminosity systems like H II galaxies and dwarf ellipticals (which are also the least metallic ones by virtue of the mass–metallicity relation). Still, it is interesting to investigate what would happen in this case. When synthesized with our  $0.2$ – $2.5 Z_{\odot}$  base (Fig. 1), these

extremely metal-poor galaxies are modelled predominantly with the  $0.2 Z_{\odot}$  components, as intuitively expected. Moreover, the mismatch in metallicity introduces non-negligible biases in other properties, like masses, mean ages and extinction ( $M_{\star}$ , for instance, is systematically underestimated by 0.3 dex). Similar problems should be encountered when modelling systems with  $Z > 2.5 Z_{\odot}$ . These results serve as a reminder that our base spans a wide but finite range in stellar metallicity, and that extrapolating these limits has an impact on the derived physical properties. While there is no straightforward a priori diagnostic of which galaxies violate these limits, in general, one should be suspicious of objects with mean  $Z_{\star}$  too close to the base limits.

### 2.4 Summary of the simulations

Summarizing this theoretical study, we have performed simulations designed to evaluate the accuracy of our spectral synthesis method. The simulations mimic inasmuch as possible the wavelength range, spectral resolution, error spectrum and S/N of the actual SDSS data studied below. Several physically motivated combinations of the synthesis parameters were investigated to establish their precision at different S/N. Table 1 summarizes the uncertainties in these quantities and a few additional ones not explicitly mentioned above. In what follows, we focus on five parameters: stellar mass, velocity dispersion, extinction, mean stellar age and metallicity, all of which were found to be adequately recovered.

This exercise demonstrates that we are capable of producing reliable estimates of several parameters of astrophysical interest, at least in principle. We must nevertheless emphasize that this conclusion relies entirely on models and on an admittedly simplistic view of galaxies. When applying the synthesis to real galaxy spectra, a series of other effects come into play. For instance, the extinction law appropriate for each galaxy probably differs from the one used here: in our Galaxy, the ratio of total to selective extinction of stellar sources,  $R_V = A_V/E(B - V)$ , is known to depend on the line of sight (Cardelli et al. 1989; Patriarchi et al. 2001); in addition, the wavelength dependence of the attenuation of light from an extended source such as a galaxy includes the effects of scattering back into the light beam, and depends on the relative distribution of stars and dust (Witt, Thronson & Capuano 1992; Gordon, Calzetti & Witt 1997). Also, one might expect that each population of stars is affected by a distinct extinction (Charlot & Fall 2000; Panuzzo et al. 2003). Similarly, while in the evolutionary tracks adopted here the metal abundances are scaled from the solar values, non-solar abundance mixtures are known to occur in stellar systems (e.g. Trager et al. 2000a,b), not to mention uncertainties in the SSP models and the always present issue of the IMF. In short, evidence against these simplifying hypotheses abounds.

Accounting for all these effects in a consistent way is not currently feasible. We mention these caveats, not to dismiss simple models, but to highlight that all parameter uncertainties discussed above are applicable within the scope of the model. Hence, while the simulations lend confidence to the synthesis method, one might remain sceptical of its actual power. The next sections further address the reliability of the synthesis, this time from a more empirical perspective.

## 3 ANALYSIS OF A VOLUME-LIMITED GALAXY SAMPLE

In this section we apply our synthesis method to a large sample of SDSS galaxies to estimate their stellar population properties. We

also present measurements of emission-line properties, obtained from the observed minus synthetic spectra. The information provided by the synthesis of so many galaxies allows one to address a long menu of astrophysical issues related to galaxy formation and evolution. Before venturing into the exploration of such issues, however, it is important to validate the results of the synthesis by as many means as possible. Hence, the goal of the study presented below is not so much to explore the physics of galaxies but to provide an empirical test of our synthesis method. The results reported in this section are used in Sections 4 and 5 with this purpose.

### 3.1 Sample definition

The spectroscopic data used in this work were taken from the SDSS. This survey provides spectra of objects in a large wavelength range (3800–9200 Å) with mean spectral resolution  $\lambda/\Delta\lambda \sim 1800$ , taken with 3 arcsec diameter fibres. The most relevant characteristic of this survey for our study is the enormous amount of good-quality, homogeneously obtained spectra. The data analysed here were extracted from the SDSS main galaxy sample available in the Data Release 2 (DR2; Abazajian et al. 2004). This flux-limited sample consists of galaxies with reddening-corrected Petrosian  $r$ -band magnitudes  $r \leq 17.77$ , and Petrosian  $r$ -band half-light surface brightnesses  $\mu_{50} \leq 24.5$  mag arcsec<sup>-2</sup> (Strauss et al. 2002).

From the main sample, we first selected spectra with a redshift confidence  $\geq 0.35$ . Following the conclusions of Zaritsky, Zabludoff & Willick (1995), we have imposed a redshift limit of  $z > 0.05$  [trying to avoid aperture effects and biases (see e.g. Gómez et al. 2003)] and selected a volume-limited sample up to  $z = 0.1$ , corresponding to an  $r$ -band absolute magnitude limit of  $M(r) = -20.5$ . The absolute magnitudes used here are  $k$ -corrected with the help of the code provided by Blanton et al. (2003, KCORRECT v3\_2) and assuming the following cosmological parameters:  $H_0 = 70$  km s<sup>-1</sup> Mpc<sup>-1</sup>,  $\Omega_M = 0.3$  and  $\Omega_\Lambda = 0.7$ . We also restricted our sample to objects for which the observed spectra show a S/N ratio in  $g$ ,  $r$  and  $i$  bands greater than 5. These restrictions leave us with a volume-limited sample containing 50 362 galaxies, which leads to a completeness level of  $\sim 98.5$  per cent.

### 3.2 Results of the spectral synthesis

All 50 362 spectra were brought to the rest frame (using the redshifts in the SDSS data base), sampled from 3650 to 8000 Å in steps of 1 Å, corrected for Galactic extinction<sup>2</sup> using the maps given by Schlegel, Finkbeiner & Davis (1998) and the extinction law of Cardelli et al. (1989, with  $R_V = 3.1$ ), and normalized by the median flux in the 4010–4060 Å region. The S/N ratio in this spectral window spans the 5–30 range, with median value of 14. Besides the masks around the lines listed in Section 2.2.1, we exclude points with SDSS flag  $\geq 2$ , which signals bad pixels, sky residuals and other artefacts. After this pre-processing, the spectra are fed into the STARLIGHT code described in Section 2.1. On average, the synthesis is performed with  $N_\lambda = 3677$  points, after discounting the ones which are clipped by our  $\leq 3$  sigma threshold (typically 40 points) and the masked ones.

The spectral fits are generally very good, as illustrated in Figs 2 and 3. The mean value of  $\chi^2/N_\lambda$  is 0.78. In fact, this is somewhat too good, since from the simulations we expect  $\chi^2/N_\lambda \sim 0.95$ . This is a minor difference, which could be fixed by decreasing the errors in  $O_\lambda$  by  $\sim 10$  per cent. We further quantify the quality of the fits by

$\Delta_\lambda$ , the mean value of  $|O_\lambda - M_\lambda|/O_\lambda$  over all non-masked points. From the simulations we expect this alternative figure of merit to be of the order of 0.6 times the noise-to-signal ratio at  $\lambda_0$  (Table 1). This is exactly the mean value of  $\Delta_\lambda \times \text{S/N}$  in the actual fits, again indicating acceptable fits.

The total stellar masses of the galaxies were obtained from the stellar masses derived from the spectral synthesis (which correspond to the light entering the fibres) by dividing them by  $(1 - f)$ , where  $f$  is the fraction of the total galaxy luminosity in the  $z$  band outside the fibre. This approach, which neglects stellar population and extinction gradients, leads to an increase of typically 0.5 in  $\log M_*$ . We did not apply any correction to the velocity dispersion estimated by the code given that the spectral resolution of the BC03 models and the data are very similar.

We point out that we did not constrain the extinction  $A_V$  to be positive. There are several reasons for this choice: (a) Some objects may be excessively dereddened by Galactic extinction. (b) Some objects may indeed require bluer SSP spectra than those in the base. (c) The observed light may contain a scattered component, which would induce a blueing of the spectra not taken into account by the adopted pure extinction law. (d) Constraining  $A_V$  to have only positive values produces an artificial concentration of solutions at  $A_V = 0$ , an unpleasant feature in the  $A_V$  distribution. Interestingly, most of the objects for which we derive negative  $A_V$  (typically  $-0.1$  to  $-0.3$  mag) are early-type galaxies. These galaxies are dominated by old populations, and expected to contain little dust. This is consistent with the result of K03, who find negative extinction primarily in galaxies with a large  $D_n(4000)$ . The distribution of  $A_V$  for these objects, which can be selected on the basis of spectral or morphological properties, is strongly peaked around  $A_V = 0$ , so that objects with  $A_V < 0$  can be considered as consistent with having zero extinction. In any case, none of the results reported in this paper is significantly affected by this choice.

### 3.3 Emission-line measurements

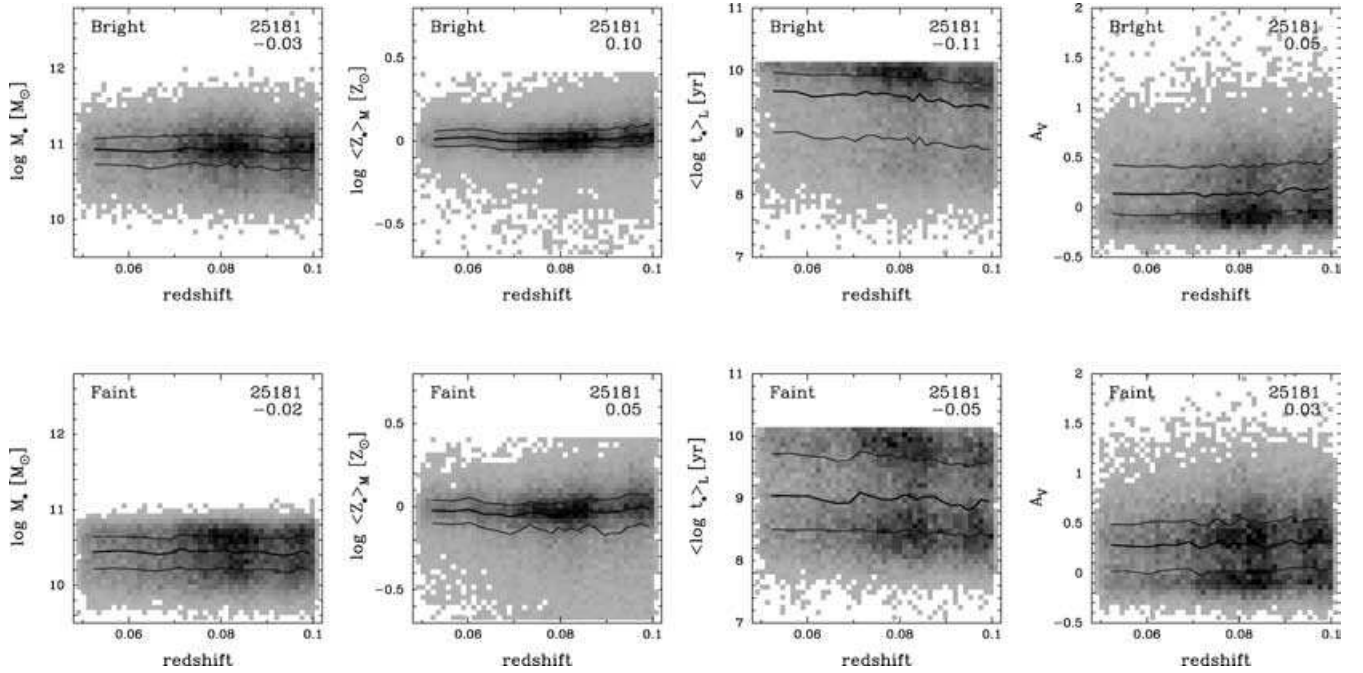
Besides providing estimates of stellar population properties, the synthesis models allow the measurement of emission lines from the ‘pure-emission’, starlight-subtracted spectra ( $O_\lambda - M_\lambda$ ). We have measured the lines of [O II]  $\lambda\lambda 3726, 3729$ , [O III]  $\lambda 4363$ , H $\beta$ , [O III]  $\lambda\lambda 4959, 5007$ , [O I]  $\lambda 6300$ , [N II]  $\lambda 6548$ , H $\alpha$ , [N II]  $\lambda 6584$  and [S II]  $\lambda\lambda 6717, 6731$ . Each line was treated as a Gaussian with three parameters: width, offset (with respect to the rest-frame central wavelength) and flux. Lines from the same ion were assumed to have the same width and offset. We have further imposed two flux ratio constraints, [O III]  $\lambda 5007$ /[O III]  $\lambda 4959 = 2.97$  and [N II]  $\lambda 6584$ /[N II]  $\lambda 6548 = 3$ . Finally, we consider a line to have significant emission if its fit presents a S/N ratio greater than 3.

In some of the following analysis, galaxies with emission lines are classified according to their position in the [O III]/H $\beta$  versus [N II]/H $\alpha$  diagram proposed by Baldwin, Phillips & Terlevich (1981) to distinguish normal star-forming galaxies from galaxies containing active galactic nuclei (AGN). We define as normal star-forming galaxies those galaxies that appear in this diagram and are below the curve defined by K03 (see also Brinchmann et al. 2004). Objects above this curve are transition objects and galaxies containing AGN.

### 3.4 Aperture bias

A point that deserves mention here is the bias that may be introduced in the analysis due to the use of small fibres to measure the galaxy spectra. This effect, known as aperture bias, may introduce a redshift dependence in the measured galaxy spectra, since the fraction of

<sup>2</sup> Unlike in the first data release, the final calibrated spectra from the DR2 are not corrected for foreground Galactic reddening.



**Figure 7.** Trends of median values (and respective quartiles) of the stellar mass, mean stellar metallicity, mean stellar age and V-band extinction as a function of redshift. Results are shown for galaxies with luminosities above (‘bright’, top panels) and below (‘faint’, bottom panels) the median luminosity of the sample. In this and the following figures, the grey-scale level represents the number of galaxies in each pixel, darker pixels having more galaxies; the upper number in the top right of each panel indicates the total number of galaxies in the plot, and the lower one is the Spearman rank correlation coefficient ( $r_S$ ).

galaxy light received by a fibre increases with increasing distance. Zaritsky et al. (1995), based on an analysis of spectra from the Las Campanas Redshift Survey, find that a lower limit on redshifts of  $z = 0.05$  minimizes the aperture bias. This effect has also been discussed by K03, Gómez et al. (2003) and Tremonti et al. (2004) for SDSS spectra. For instance, these authors found that the galaxy  $z$ -band  $M/L$  ratio and the gas-phase oxygen abundance decrease by  $\sim 0.1$  dex over the redshift range of the survey, indicating that these quantities are moderately affected by aperture bias. The redshift range considered in our study is smaller than that of the entire survey, implying that the aperture effects are expected to be even smaller in our sample.

In order to verify whether our results are significantly affected by this bias, we investigated the behaviour of some of the quantities resulting from our analysis ( $M_*$ ,  $\langle Z_* \rangle$ ,  $\langle \log t_* \rangle$  and  $A_V$ ) as a function of redshift for galaxies with luminosities above and below  $M(r) = -21.17$ , the median luminosity of our sample. We divided the sample into several redshift bins containing the same number of objects, and computed the median value and the quartiles of each quantity in each bin. Fig. 7 shows as solid lines the median values of each distribution, as well as their respective, quartiles. None of the quantities seems to be strongly affected by this bias. The largest correlation with  $z$  appears for light-weighted ages, corresponding to a change along our redshift distribution of about 0.1 and 0.26 dex for faint and bright galaxies, respectively. This is a plausible result since the light of nearby objects seen by the fibre aperture is dominated by the older stellar populations of their bulges. For the other quantities, the variations are well below 0.1 dex.

#### 4 COMPARISONS WITH THE MPA/JHU DATA BASE

The SDSS data base has been explored by several groups, using different approaches and techniques. The MPA/JHU group has re-

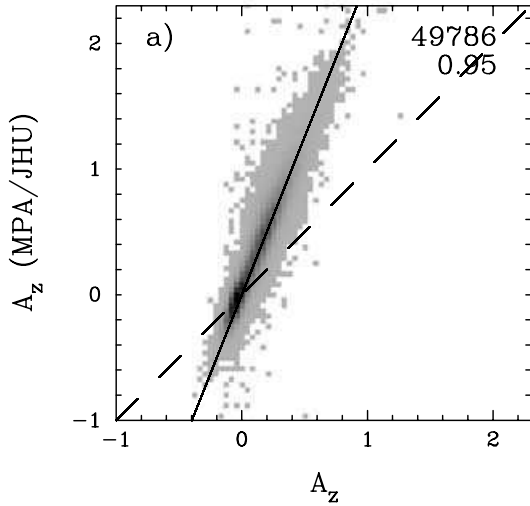
cently publicly released catalogues<sup>3</sup> of derived physical properties for 211 894 SDSS galaxies, including 33 589 narrow-line AGN (K03; see also Brinchmann et al. 2004). These catalogues are based on the K03 method to infer the star formation histories, dust attenuation and stellar masses of galaxies from the simultaneous analysis of the 4000-Å break strength,  $D_n(4000)$ , and the Balmer line absorption index  $H\delta_A$ . These two indices are used to constrain the mean stellar ages of galaxies and the fractional stellar mass formed in bursts over the past few Gyr, and a comparison with broad-band photometry then allows the extinction and stellar masses to be estimated.

The MPA/JHU catalogues provide very useful benchmarks for similar studies. In this section we compare the values of some of the parameters from these catalogues with our own estimates. Catalogues of galaxy properties obtained with our synthesis method will be made available in due course. Besides comparing directly measurements of physical quantities, our aim is also to highlight the differences that may appear due to the use of different methods and procedures.

##### 4.1 Stellar extinction

The MPA/JHU group estimates the  $z$ -band stellar extinction  $A_z$  through the difference between model and measured colours, assuming an attenuation curve proportional to  $\lambda^{-0.7}$ . In our case, the extinction  $A_V$  is derived directly from the spectral fitting, carried out with the Milky Way extinction law (Cardelli et al. 1989, cf. Section 2.1), for which  $A_z = 0.4849 A_V$  assuming  $\lambda_z = 8931 \text{ \AA}$ . Fig. 8 shows that these two independent estimates are very strongly and linearly correlated, with a Spearman rank correlation coefficient  $r_S = 0.95$ . However, the values of  $A_z$  reported by the MPA/JHU group

<sup>3</sup> Available at <http://www.mpa-garching.mpg.de/SDSS/>



**Figure 8.** Comparison of the stellar extinction in the  $z$  band determined by the MPA/JHU group with that estimated through the method described in this paper. The solid line shows a  $y = ax$  linear fit to the data, while the dashed one is the identity line ( $y = x$ ).

(column 17 of their stellar mass catalogue) are systematically larger than our values:  $A_z(\text{MPA/JHU}) \simeq 2.51 A_z(\text{this work})$  in the median.

This discrepancy is only apparent. The Galactic extinction law is substantially harder than  $\lambda^{-0.7}$ . One thus expects to need less extinction when modelling a given galaxy with the former law than with the latter. This was confirmed by STARLIGHT fits to a subset of SDSS galaxies using a  $\lambda^{-0.7}$  law, which yield a value of  $A_V$  typically 1.77 times larger than those obtained with the Cardelli et al. (1989) law. Since the  $A_z/A_V$  conversion factors are 0.7030 and 0.4849 for the  $\lambda^{-0.7}$  and Cardelli et al. laws, one finds that the  $A_z$  values obtained for the two laws should differ by a factor of  $1.77 \times 0.7030/0.4849 = 2.57$ . This is very close to the empirically derived factor of 2.51 (Fig. 8). We thus conclude that there are no substantial differences between the MPA/JHU and our estimates of the stellar extinction other than those implied by the differences in the reddening laws adopted in the two studies.

## 4.2 Stellar masses

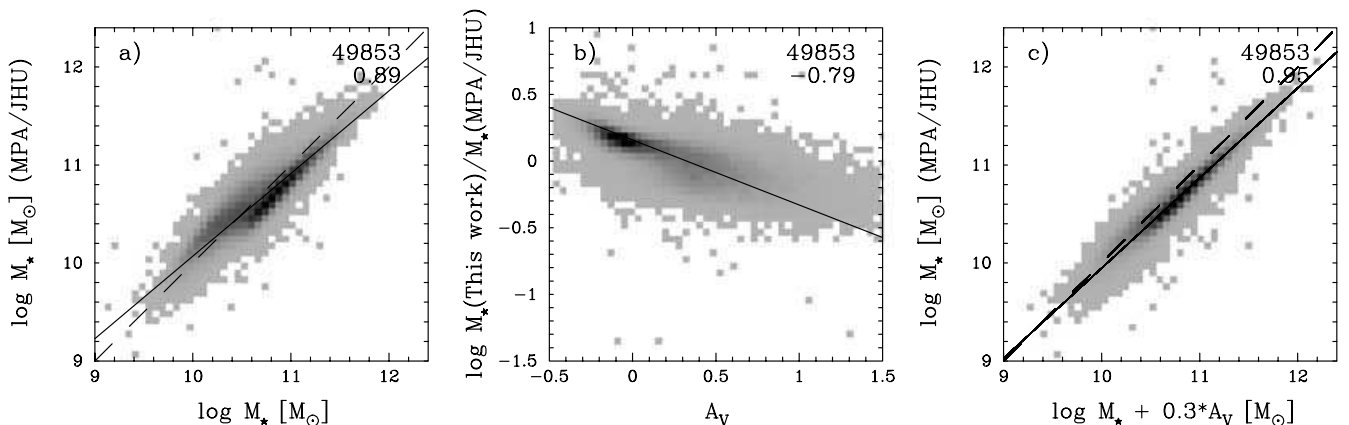
Fig. 9(a) compares our results for the total stellar masses to the MPA/JHU extinction-corrected stellar masses (column 9 of their stellar mass catalogue). The two estimates of  $M_*$  correlate very well, with  $r_S = 0.89$ . The quantitative agreement is also good, with a median difference of just 0.1 dex. This small offset cannot be attributed to the different IMFs employed in the two studies (Chabrier 2003 here and Kroupa 2001 in K03), since, as illustrated in fig. 4 of BC03, these two IMFs yield practically identical  $M/L$  ratios. Instead, this offset seems to be due to a subtle technicality. Whereas we adopt the  $M/L$  ratio of the best  $\chi^2$  model, the MPA/JHU group derives  $M/L$  by comparing the observed values of the  $D_n(4000)$  and  $H\delta_A$  indices with a library of 32 000 models. Each model is then weighted by its likelihood, and a probability distribution for  $M/L$  is computed. The MPA/JHU mass is the median of this distribution, which is not necessarily the same as the best  $\chi^2$  value. In fact, the stellar mass catalogue of Brinchmann et al. (2004) lists in its column 8 the best  $\chi^2$  masses, which are systematically larger than the median ones by  $\sim 0.1$  dex, identical to the median difference identified above.

Part of the scatter in Fig. 9(a) can be attributed to the different extinction laws. Extinction contributes  $+0.4 A_{\lambda_0}$  to the estimated  $\log M_*$ . As discussed above, using a  $\lambda^{-0.7}$  law in the synthesis yields values of  $A_V$  that are 1.77 times larger than using the Cardelli et al. (1989) law. Furthermore, the mass-to-light ratios obtained with the two laws are very similar. From this we expect that

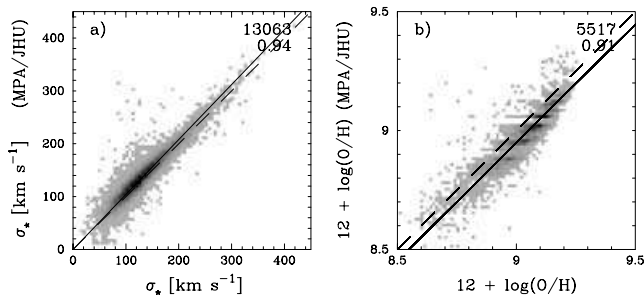
$$\log M_*(\text{this work})/M_*(\text{MPA/JHU}) \approx -0.3 A_V(\text{this work}),$$

in good agreement with the observed relation (Fig. 9b). Correcting for this effect by adding  $0.3 A_V$  to our mass estimates indeed produces a better correlation, with  $r_S = 0.95$ , as illustrated in Fig. 9(c).

Overall, we conclude that the two mass estimates agree to within 0.4 dex. This level of agreement is similar to that recently found by Drory, Bender & 2004, (2004) in their comparison between the MPA/JHU masses and estimates based on SDSS plus 2MASS photometry. It is quite remarkable that, despite the substantial differences between our approaches and the underlying assumptions, the estimated stellar masses are so similar over a wide range of masses. On the other hand, this may not be so surprising given that the MPA/JHU group estimates of physical properties are ultimately based on a comparison of observed indices with an extensive library



**Figure 9.** (a) Comparison of the stellar mass determined by the MPA/JHU group with our estimate. The solid line shows a  $y = ax + b$  linear fit to the data, while the dashed one is the identity line ( $y = x$ ). (b) Relation of the ratio between our and the MPA/JHU masses and our estimate of the  $V$ -band stellar extinction. (c) Comparison of the stellar masses after correcting for differences induced by the different extinction laws.



**Figure 10.** Comparison of (a) the velocity dispersions and (b) the oxygen abundances determined by the MPA/JHU group with our estimates. The solid line shows a  $y = ax + b$  linear fit to the data, while the dashed one is the identity line ( $y = x$ ).

of galaxy spectra constructed out of the BC03 evolutionary synthesis models (K03). Since this library spans a wide of metallicities and star formation histories, the agreement between our estimates and those of the MPA/JHU group may be simply indicating that the  $N_* = 45$  SSPs from BC03 used in our spectral synthesis span a similar parameter space to that covered by the K03 models.

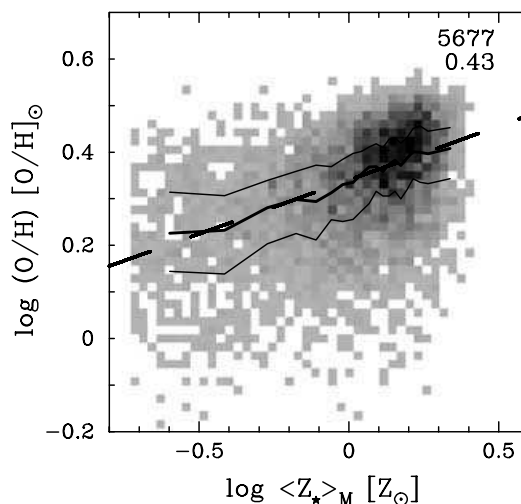
### 4.3 Velocity dispersion

We use the subsample of galaxies with active nuclei to compare our measurements of absorption-line broadening due to galaxy velocity dispersion and/or rotation,  $\sigma_*$ , with those of the MPA/JHU group, since they list this quantity only in their AGN catalogue (in column 16). The comparison is displayed in Fig. 10(a). The Spearman correlation coefficient in this case is  $r_s = 0.91$  and the median of the difference between the two estimates is just  $9 \text{ km s}^{-1}$ , indicating an excellent agreement between the two studies.

### 4.4 Emission lines and nebular metallicities

Brinchmann et al. (2004) also provide, in their emission-line catalogue, emission-line fluxes and equivalent widths, which can be compared with our own measurements. In both studies the emission lines are measured after subtracting from the observed spectrum a model spectrum representing the stellar emission. In our case this is done with our synthesized spectra. The MPA/JHU group adopts a similar approach (see Tremonti et al. 2004 for a brief description) by fitting the observed continuum with BC03 models. They adopt, however, a single metallicity model and a different extinction law. We have compared the fluxes and equivalent widths of  $H\alpha$ ,  $[\text{N II}] \lambda 6584$ ,  $[\text{O II}] \lambda 3727$ ,  $H\beta$  and  $[\text{O III}] \lambda 5007$  as measured by our code and that obtained by the MPA/JHU group. We do not find any significant difference between these values; the largest discrepancy ( $\sim 5$  per cent) was found for the equivalent widths of  $H\alpha$  and  $[\text{N II}]$ , probably due to different estimates of the continuum level and the associated underlying stellar absorption. Our emission-line measurements are also in good agreement with those in Stasińska et al. (2004), who fit the Balmer lines with emission and absorption components, instead of subtracting a starlight model.

In Fig. 10(b) we plot our estimates of the nebular oxygen abundance against those obtained by the MPA/JHU group (catalogue of gas-phase metallicities, median values; see Tremonti et al. 2004). In order not to introduce any bias due to the use of different indicators of the oxygen abundance, we have estimated O/H using the calibration of the  $([\text{O II}] \lambda 3726, 3729 + [\text{O III}] \lambda 4959, 5007)/H\beta$  ratio given by Tremonti et al. (2004) in their equation (1). This calibration is



**Figure 11.** Nebular oxygen abundance versus mass-weighted stellar metallicity, both in solar units, for normal star-forming galaxies in our sample. The median values and quartiles in bins of same number of objects are shown as thin solid lines. The dashed line is a robust fit for the relation.

based on simultaneous fits of the most prominent emission lines with a model designed for the interpretation of integrated galaxy spectra (Charlot & Longhetti 2001). The oxygen abundances estimated in such a way differ by just  $\sim 0.05$  dex, as shown in Fig. 10(b).

Overall, we conclude that our spectral synthesis method yields estimates of physical parameters in good agreement with those obtained by the MPA/JHU group, considering the important differences in approach and underlying assumptions.

## 5 EMPIRICAL RELATIONS

Yet another way to assess the validity of physical properties derived through a spectral synthesis analysis is to investigate whether this method yields astrophysically reasonable results. In this section we follow this empirical line of reasoning by comparing some results obtained from our synthesis of SDSS galaxies (which excludes emission lines) with those obtained from a direct analysis of the emission lines. Our aim is to demonstrate that our synthesis results do indeed make sense. A more detailed discussion of most of the points below will be presented in other papers of this series.

### 5.1 Stellar and nebular metallicities

Our spectral synthesis approach yields estimates of the mean metallicity of the stars in a galaxy,  $\langle Z_* \rangle$ . The analysis of emission lines, on the other hand, gives estimates of the present-day abundances in the warm interstellar medium. Although stellar and nebular metallicities are not expected to be equal, it is reasonable to expect that they should roughly scale with each other.

Fig. 11 shows the correlation between mass-weighted stellar metallicities and the nebular oxygen abundance (computed as in Section 4.4), both in solar units,<sup>4</sup> for our sample of normal star-forming galaxies. A correlation is clearly seen, although with large scatter ( $r_s = 0.42$ ). Galaxies with large stellar metallicities also have large nebular oxygen abundances; galaxies with low stellar

<sup>4</sup> The solar unit adopted for the nebular oxygen abundance is  $12 + \log(\text{O}/\text{H})_\odot = 8.69$  (Allende Prieto, Lambert & Asplund 2001).

abundances tend to have smaller abundances. The observed scatter is qualitatively expected due to variations in enrichment histories among galaxies. A robust linear fitting gives the following relation:

$$\log\left(\frac{\text{O}/\text{H}}{0.00049}\right) = 0.34 + 0.23 \log\left(\frac{\langle Z_* \rangle_M}{0.02}\right), \quad (7)$$

with a dispersion of 0.08 dex. Notice that in this expression both stellar and nebular metallicities are normalized to solar units,  $Z_\odot = 0.02$  and  $(\text{O}/\text{H})_\odot = 0.00049$ , respectively.

Nebular and stellar metallicities are estimated through completely different and independent methods, so the correlation depicted in Fig. 11 provides an a posteriori empirical validation for the stellar metallicity derived by the spectral synthesis. The possibility to estimate stellar metallicities for so many galaxies is one of the major virtues of spectral synthesis, as it opens an important window to study the chemical evolution of galaxies and of the Universe as a whole (Panter et al. 2004; Sodré et al., in preparation).

## 5.2 Stellar and nebular extinctions

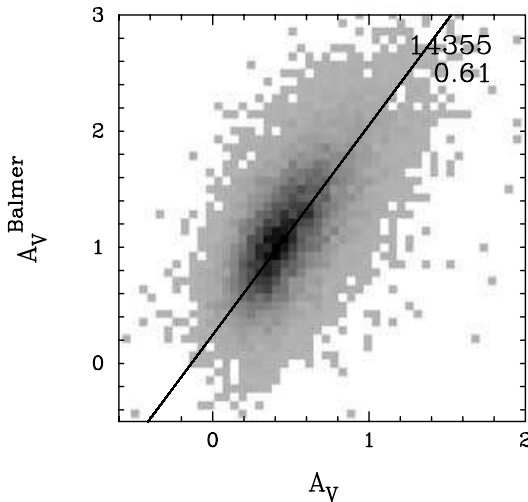
The stellar extinction in the  $V$  band is one of the products of our STARLIGHT code. A more traditional and completely independent method to evaluate the extinction consists of comparing the observed  $\text{H}\alpha/\text{H}\beta$  Balmer decrement to the theoretical value. The intrinsic value of  $F(\text{H}\alpha)/F(\text{H}\beta)$  is not very sensitive to the physical conditions of the gas, ranging from 3.03 for a gas temperature of 5000 K to 2.74 at 20 000 K (Osterbrock 1989). Adopting a value of 2.86 for this ratio, the ‘Balmer extinction’ (Stasińska et al. 2004) is given by

$$A_V^{\text{Balmer}} = 6.31 \log\left[\frac{F(\text{H}\alpha)/F(\text{H}\beta)}{2.86}\right], \quad (8)$$

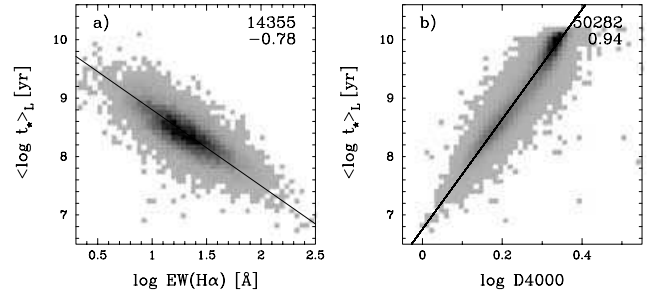
where the coefficient 6.31 comes from assuming the Cardelli et al. (1989) extinction curve.

Fig. 12 presents a comparison between the stellar  $A_V$  and  $A_V^{\text{Balmer}}$ . These two extinctions are determined in completely independent ways, and yet our results show that they are closely linked, with  $r_S = 0.61$ . A linear bisector fitting yields

$$A_V^{\text{Balmer}} = 0.24 + 1.81 A_V. \quad (9)$$



**Figure 12.** Relation between the nebular ( $A_V^{\text{Balmer}}$ ) and stellar ( $A_V$ ) extinctions for our sample of normal star-forming galaxies. The solid line is a robust fit for the relation.



**Figure 13.** (a) Equivalent width of  $\text{H}\alpha$  versus the light-weighted mean stellar age for normal emission-line galaxies in our sample. (b) Relation between the D4000 index and the light-weighted mean stellar age. The solid lines are robust fits for the relations.

Note that the angular coefficient in this relation indicates that nebular photons experience roughly twice as much extinction of the starlight. This ‘differential extinction’ is in very good qualitative and quantitative agreement with empirical studies (Fanelli, O’Connell & Thuan 1988; Calzetti, Kinney & Storchi-Bergmann 1994; Gordon et al. 1997; Mas-Hesse & Kunth 1999). We shall explore the implications of this result for the intrinsic colours of star-forming galaxies in another paper of this series.

## 5.3 Relations with mean stellar age

The equivalent width (EW) of  $\text{H}\alpha$  is related to the ratio of present to past star formation rate of a galaxy (e.g. Kennicutt 1998). It is thus expected to be smaller for older galaxies. Fig. 13(a) shows the relation between  $\text{EW}(\text{H}\alpha)$  and the mean light-weighted stellar age obtained by our spectral synthesis. The anticorrelation, which has  $r_S = -0.78$ , is evident. From this plot we can derive an empirical relation that can be used to estimate  $\langle \log t_* \rangle_L$  (for  $\lambda_0 = 4020 \text{ \AA}$ ) of star-forming galaxies through the measurement of  $\text{EW}(\text{H}\alpha)$ :

$$\langle \log t_* \rangle_L = 10.10 - 1.30 \log \text{EW}(\text{H}\alpha), \quad (10)$$

for  $t_*$  in yr and  $\text{EW}(\text{H}\alpha)$  in  $\text{\AA}$ . It is worth stressing that these two quantities are obtained independently, since the spectral synthesis does not include emission lines.

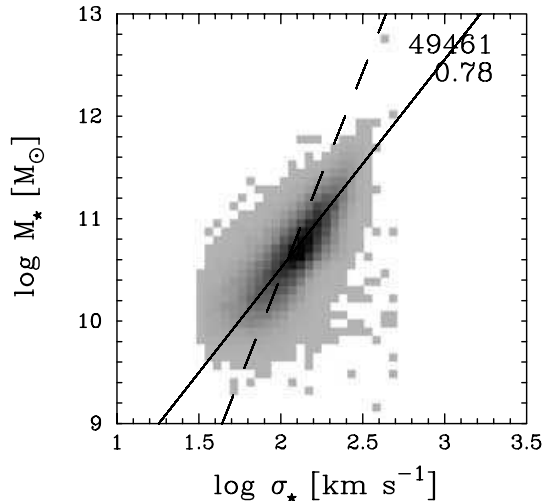
Another quantity that is considered a good age indicator, even for galaxies without emission lines, is the 4000- $\text{\AA}$  break, D4000. We measured this index following Bruzual (1983), who define D4000 as the ratio between the average value of  $F_V$  in the 4050–4250 and 3750–3950  $\text{\AA}$  bands. The relation between  $\langle \log t_* \rangle_L$  and D4000 is shown in Fig. 13(b). Note that the concentration of points at the high-age end reflects the upper age limit of the base adopted here, 13 Gyr (cf. Section 2.2.1). The correlation is very strong ( $r_S = 0.94$ ), showing that indeed D4000 can be used to estimate empirically mean light-weighted galaxy ages, despite its metallicity dependence for very old stellar populations (older than 1 Gyr, as shown by K03). From the tight relation between  $\langle \log t_* \rangle_L$  and D4000, we derive the following empirical relation:

$$\langle \log t_* \rangle_L = 6.76 + 9.41 \log \text{D4000}, \quad (11)$$

where  $t_*$  is in yr. This robust fit reproduces  $\langle \log t_* \rangle_L$  to within an rms dispersion of 0.15 dex.

## 5.4 $M_*$ – $\sigma_*$ relation

Fig. 14 shows the relation between stellar mass and  $\sigma_*$ , obtained from the synthesis. The relation is quite good, with  $r_S = 0.79$ . The



**Figure 14.** Mass–velocity dispersion relation for our sample. The solid line is a robust fit for the data, while the dashed one is a fit assuming  $M_* \propto \sigma_*^4$ .

solid line displayed in the figure is

$$\log M_* = 6.44 + 2.04 \log \sigma_*, \quad (12)$$

for  $M_*$  in  $M_\odot$  and  $\sigma_*$  in  $\text{km s}^{-1}$ , obtained with a bisector fitting. The figure also shows as a dashed line a fit assuming  $M_* \propto \sigma_*^4$ , expected from the virial theorem under the (unrealistic) assumption of constant mass surface density. In both cases we have excluded from the fit galaxies with  $\sigma_* < 35 \text{ km s}^{-1}$ , which corresponds to less than half the spectral resolution of both data and models.

This is another relation that is expected a priori if we have in mind the Faber–Jackson relation for ellipticals and the Tully–Fisher relation for spirals. For early-type galaxies,  $\sigma_*$  is a measure of the central velocity dispersion, which is directly linked to the gravitational potential depth and, through the virial theorem, to galactic mass. For late-type systems,  $\sigma_*$  has contributions of isotropic motions in the bulges, as well as of the rotation of the discs, and is also expected to be related with galactic mass. Another aspect that it is interesting to point out in Fig. 14 is that the dispersion in the  $M_*$ – $\sigma_*$  relation decreases as we go from low-luminosity, rotation-dominated systems, for which the values of  $\sigma_*$  depend on galaxy inclination and bulge-to-disc ratio, to high-luminosity, mostly early-type systems, which obey a much more regular (and steeper) relation between  $\sigma_*$  and  $M_*$ .

This relation, between a quantity that is not directly linked to the synthesis,  $\sigma_*$ , and another one that is a product of our synthesis,  $M_*$ , is yet another indication that the results of our STARLIGHT code do make sense.

## 6 SUMMARY

We have developed and tested a method to fit galaxy spectra with a combination of spectra of individual simple stellar populations generated with state-of-the-art evolutionary synthesis models. The main goal of this investigation was to examine the reliability of physical properties derived in this way. This goal was pursued by three different means: simulations, comparison with independent studies, and analysis of empirical results. Our main results can be summarized as follows:

(i) Simulations tailored to match the characteristics of SDSS spectra show that the individual SSP strengths, encoded in the popu-

lation vector  $\mathbf{x}$ , are subjected to large uncertainties, but robust results can be obtained by compressing  $\mathbf{x}$  into coarser but useful indices. In particular, physically motivated indices such as mean stellar ages and metallicities are found to be well recovered by spectral synthesis even for relatively noisy spectra. Stellar masses, velocity dispersion and extinction are also found to be accurately retrieved.

(ii) We have applied our STARLIGHT code to a volume-limited sample of over 50 000 galaxies from the SDSS Data Release 2. The spectral fits are generally very good, and allow accurate measurements of emission lines from the starlight-subtracted spectrum. Catalogues of physical and emission-line properties derived for this sample will be made available in due course. We also report that work is under way to produce a flexible, user-friendly and publicly available version of STARLIGHT.

(iii) We have compared our results to those obtained by the MPA/JHU group (K03; Brinchmann et al. 2004) with a different method to characterize the stellar populations of SDSS galaxies. The stellar extinctions and masses derived in these two studies are very strongly correlated. Furthermore, differences in the values of  $A_V$  and  $M_*$  are found to be mostly due to the differences in the model ingredients (extinction law). Our estimates of stellar velocity dispersions and emission-line properties are also in good agreement with those of the MPA/JHU group.

(iv) The confidence in the method is further strengthened by several empirical correlations between synthesis results and independent quantities. We find strong correlations between stellar and nebular metallicities, stellar and nebular extinctions, mean stellar age and the equivalent width of  $\text{H}\alpha$ , mean stellar age and the 4000-Å break, stellar mass and velocity dispersion. These are all astrophysically reasonable results, which reinforce the conclusion that spectral synthesis is capable of producing reliable estimates of the physical properties of galaxies.

Overall, these results validate spectral synthesis as a powerful tool to study the history of galaxies. Other papers in this series will take advantage of this tool to address issues regarding aspects of galaxy formation and evolution.

## ACKNOWLEDGMENTS

We thank the organizers of the Guillermo Haro Workshop 2004 at the Instituto Nacional de Astronomía, Óptica y Electrónica (Puebla, Mexico) for having allowed us to work in a very pleasant and stimulating environment, and we thank the participants for many useful discussions. We are also in debt to the anonymous referee for her/his comments and helpful suggestions. Partial support from CNPq, FAPESP and the France–Brazil PICS programme are also acknowledged. Last but not least, we wish to thank G. Bruzual, S. Charlot and the SDSS team for their dedication to projects which made the present work possible.

The Sloan Digital Sky Survey is a joint project of the University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, the Johns Hopkins University, the Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Princeton University, the United States Naval Observatory, and the University of Washington. Funding for the project has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the US Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society.

## REFERENCES

- Abazajian K. et al., 2003, *AJ*, 126, 2081  
 Abazajian K. et al., 2004, *AJ*, 128, 502  
 Allende Prieto C., Lambert D. L., Asplund M., 2001, *ApJ*, 556, L63  
 Arimoto N., Yoshii Y., 1987, *A&A*, 173, 23  
 Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, 93, 5  
 Bertone E., Buzzoni A., Rodríguez-Merino L. H., Chávez M., 2004, *Mem. Soc. Astron. Ital.*, 75, 158  
 Bica E., 1988, *A&A*, 195, 76  
 Blanton M. R. et al., 2003, *AJ*, 125, 2348  
 Bressan A., Chiosi C., Tantaló R., 1996, *A&A*, 311, 425  
 Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, *MNRAS*, 351, 1151  
 Bruzual G., 1983, *ApJ*, 273, 105  
 Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000 (BC03)  
 Calzetti D., Kinney A. L., Storchi-Bergmann T., 1994, *ApJ*, 429, 582  
 Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, 345, 245  
 Cardiel N., Gorgas J., Sánchez-Blázquez P., Cenarro A. J., Pedraz S., Bruzual G., Klement J., 2003, *A&A*, 409, 511  
 Chabrier G., 2003, *PASP*, 115, 763  
 Charlot S., Fall S. M., 2000, *ApJ*, 539, 718  
 Charlot S., Longhetti M., 2001, *MNRAS*, 323, 887  
 Cid Fernandes R., Sodr e L., Schmitt H. R., Leão J. R. S., 2001, *MNRAS*, 325, 60  
 Cid Fernandes R., Leão J., Lacerda R. R., 2003, *MNRAS*, 340, 29  
 Cid Fernandes R., Gu Q., Melnick K., Terlevich E., Terlevich R., Kunth D., Rodrigues Lacerda R., Joguet B., 2004, *MNRAS*, 355, 273 (CF04)  
 Cid Fernandes R., González Delgado R., Storchi-Bergmann T., Pires Martins L., Schmitt H., 2005, *MNRAS*, 356, 270  
 Doran M. Müller C. M., 2004, *J. Cosmol. Astropart. Phys.*, 09JCAP003  
 Drory N., Bender R., Hopp U., 2004, *ApJ*, 616, L103  
 Faber S. M., 1972, *A&A*, 20, 361  
 Fanelli M. N., O'Connell R. W., Thuan T. X., 1988, *ApJ*, 334, 665  
 Fioc M., Rocca-Volmerange B., 1997, *A&A*, 326, 950  
 Garcia-Rissmann A., Vega L. R., Asari N. V., Cid Fernandes R., González Delgado R. M., Storchi-Bergman T., 2005, *MNRAS*, in press  
 Gómez P. L. et al., 2003, *ApJ*, 584, 210  
 González Delgado R. M., Cerviño M., Martins L. P., Leitherer C., Hauschildt P. H., 2005, *MNRAS*, 357, 945  
 Gordon K. D., Calzetti D., Witt A. N., 1997, *ApJ*, 487, 625  
 Guiderdoni B., Rocca-Volmerange B., 1987, *A&A*, 186, 1  
 Heavens A. F., Jimenez R., Lahav O., 2000, *MNRAS*, 317, 965  
 Heavens A., Panter B., Jimenez R., Dunlop J., 2004, *Nat.*, 428, 625  
 Jimenez R., MacDonald J., Dunlop J. S., Padoan P., Peacock J. A., 2004, *MNRAS*, 349, 240  
 Kauffmann G. et al., 2003, *MNRAS*, 341, 33 (K03)  
 Kennicutt R. C. Jr, 1998, *ARA&A*, 36, 189  
 Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, *ApJ*, 556, 121  
 Kroupa P., 2001, *MNRAS*, 322, 231  
 Le Borgne J.-F. et al., 2003, *A&A*, 402, 433  
 Le Borgne D., Rocca-Volmerange B., Prugniel P., Lançon A., Fioc M., Soubiran C., 2004, *A&A*, 425, 881  
 Leitherer C., Robert C., Heckman T. M., 1995, *ApJS*, 99, 173  
 Loveday J., Peterson B. A., Maddox S. J., Efstathiou G., 1996, *ApJS*, 107, 201  
 MacKay D. J. C., 2003, *Information Theory, Inference and Learning Algorithms*. Cambridge Univ. Press, Cambridge  
 Mas-Hesse J. M., Kunth D., 1991, *A&AS*, 88, 399  
 Mas-Hesse J. M., Kunth D., 1999, *A&AS*, 349, 765  
 Morgan W. W., 1956, *PASP*, 68, 509  
 Moultaqa J., Pelat D., 2000, *MNRAS*, 314, 409  
 Moultaqa J., Boisson C., Joly M., Pelat D., 2004, *A&A*, 420, 459  
 Osterbrock D. E., 1989, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei*. University Science Books, Mill Valley CA  
 Panter B., Heavens A. F., Jimenez R., 2004, *MNRAS*, 355, 764  
 Panuzzo P., Bressan A., Granato G. L., Silva L., Danese L., 2003, *A&A*, 409, 99  
 Patriarchi P., Morbidelli L., Perinotto M., Barbaro G., 2001, *A&A*, 372, 644  
 Pelat D., 1997, *MNRAS*, 284, 365  
 Prugniel P., Soubiran C., 2001, *A&A*, 369, 1048  
 Reichardt C., Jimenez R., Heavens A. F., 2001, *MNRAS*, 327, 849  
 Renzini A., Buzzoni A., 1986, in Chiosi C., Renzini A., eds, *Spectral Evolution of Galaxies*. Riedel, Dordrecht, p. 213  
 Ronen S., Aragon-Salamanca A., Lahav O., 1999, *MNRAS*, 303, 284  
 Schaerer D., Vacca W. D., 1998, *ApJ*, 497, 618  
 Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525  
 Schmidt A. A., Copetti M. V. F., Alloin D., Jablonka P., 1991, *MNRAS*, 249, 766  
 Spinrad H., Taylor B. J., 1972, *ApJ*, 171, 397  
 Stasińska G., Mateus A., Sodr e L., Szczerba R., 2004, *A&A*, 420, 475  
 Stoughton C. et al., 2002, *AJ*, 123, 485  
 Strauss M. A. et al., 2002, *AJ*, 124, 1810  
 Tinsley B. M., 1968, *ApJ*, 151, 547  
 Trager S. C., Faber S. M., Worthey G., Jesús González J., 2000a, *AJ*, 119, 1645  
 Trager S. C., Faber S. M., Worthey G., Jesús González J., 2000b, *AJ*, 120, 165  
 Tremonti C., 2003, PhD thesis, Johns Hopkins Univ.  
 Tremonti C. A. et al., 2004, *ApJ*, 613, 898  
 Vazdekis A., 1999, *ApJ*, 513, 224  
 Vazdekis A., Arimoto N., 1999, *ApJ*, 525, 144  
 Witt A. N., Thronson H. A., Capuano J. M., 1992, *ApJ*, 393, 611  
 Wood D. B., 1966, *ApJ*, 145, 36  
 Worthey G., 1994, *ApJS*, 94, 687  
 York D. G. et al., 2000, *AJ*, 120, 1579  
 Zaritsky D., Zabludoff A. I., Willick J. A., 1995, *AJ*, 110, 1602

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.



# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)