



**Observatório
Nacional**

Ministério da Ciência, Tecnologia e Inovação
Divisão de Programas de Pós-Graduação
Astronomia e Astrofísica

Morfologia automatizada de galáxias com altas taxas de formação estelar

Didier Curty

Orientadores:

Prof. Dr. Eduardo Telles
Prof. Dr. François Cuisinier

Tese apresentada como requisito para a obtenção do título de

Doutor em Astronomia

Rio de Janeiro — Março de 2012

A Ambinintsoa, Émilie e Adam.



Toutefois il se peut faire que je me trompe, et ce n'est peut-être qu'un peu de cuivre et de verre que je prends pour de l'or et des diamants. Je sais combien nous sommes

sujets à nous méprendre en ce qui nous touche, et combien aussi les jugements de nos amis nous doivent être suspects, lorsqu'ils sont en notre faveur.

[...]

Mais pourcequ'alors je désirois vaquer seulement à la recherche de la vérité, je pensai qu'il falloit que je fisse tout le contraire, et que je rejetasse comme absolument faux tout ce en quoi je pourrois imaginer le moindre doute, afin de voir s'il ne resteroit point après cela quelque chose en ma créance qui fut entièrement indubitable.



RENÉ DESCARTES

Discours de la méthode (1637)

pour bien conduire sa raison, et chercher la vérité dans les sciences



Prefácio

Um doutorado é muito mais que alguns anos de pesquisa. É um pedaço de vida.

Uma tarde de fim de verão de 2007, cheguei ao Brasil com a minha recente esposa. Não conhecíamos nada da sociedade brasileira, não falávamos nenhuma palavra de português. Hoje, temos dois filhos brasileiros, uma experiência cultural inestimável, o conhecimento de uma nova língua (que sei, ainda é deficiente; agradeço ao leitor por sua indulgência) e a cabeça cheia de lembranças deliciosas.

Em mais de quatro anos, aprendi muito de Astronomia, mais também muito do Brasil e dos brasileiros. Oh, dizer que todos os dias foram uma sinecura seria uma mentira! Houve dificuldades e decepções. Mas houve, sobretudo, satisfações, alegria, encantamento. Todos aqueles que encontrei aqui, uma vez, uma hora, ou que me acompanharam ao longo desta tese, todos me permitiram descobrir um pouco do Brasil: o samba aqui, o futebol ali, as praias, o feijão, a bagunça, o calor, ou a espontaneidade carioca.

Tenho, nesse momento, um pensamento especial para o Professor François Cuisinier, que orientou minha tese, mas que uma leucemia não lhe deixou ver concluir. As escolhas de vida devem muito aos acasos dos encontros. A primeira vez que encontrei François, em frente ao instituto de Astronomia da universidade de Basileia (Suíça), ele já me falou do Brasil e da possibilidade de um doutorado. Dois anos depois, aterrissei no Galeão. Sem François, eu não estaria aqui; sou-lhe grato pelos cinco anos passados. Seu sorriso e sua risada ruidosa residem nas minhas lembranças brasileiras.

Vilma **Thanks**
Ana Cristina e Paulo
Carlos **O meu pai** **Beethoven** **Pieter** **Roland**
Danke **Denise** **Gustavo**
Émilie **Neide** **Maria Isela**
Chico **Vladimir**
Rodrigo **Alberto** **O Observatório Nacional**
Merci **Misaotra**
Orlando **A minha família**
Aldinêz **Eduardo**
Ambinintsoa **Gustavo** **A minha mai**
Jamille
Nobar **Júlio** **Anderson**
Vladimir **João**
Toda a Vila Pacheco de Catete **Adam**
Flávia **Obrigado**
Fernando **Pensées** **Bruno**
Henri **João Victor** **Manu**
François **Ernandes** **A CAPES**
Agradecimentos **Daniel**

Resumo

A presente tese trata principalmente de fotometria superficial, tendo como alvo a manipulação de grande quantidade de dados. Implementamos os nossos próprios processos, inteiramente automatizados, de redução de dados e de cálculo de parâmetros fotométricos. Uma grande parte desse trabalho consiste na descrição técnica dos métodos desenvolvidos.

Aplicamos o nosso procedimento ao estudo da morfologia de galáxias em um catálogo de 2253 objetos, cujo tipo de Hubble era previamente conhecido. As quantidades fotométricas foram computadas a partir da determinação da “isofota de Petrosian”, uma abertura que reúne as vantagens de isofotas e de um raio característico. Calculamos os índices morfológicos habituais: o parâmetro de Sérsic n , o índice de concentração C , o índice de assimetria A , o índice de *clumpiness* S , o coeficiente de Gini G e o índice M_{20} . Tentamos melhorar os processos de computação, por exemplo pela utilização sistemática de aberturas elípticas.

Definimos o índice de anelipticidade α como uma avaliação da deformação da isofota externa da galáxia, e o índice de granulação γ como a inclinação do espectro formado pelos momentos radiais de segunda ordem dos coeficientes de *wavelet*; γ pode ser interpretado como um parâmetro característico de uma galáxia que mede regiões de formação estelar distantes do seu centro.

As correlações entre os índices nos permitem distinguir duas famílias paramétricas: n , C e G fornecem uma medida da concentração de luz; enquanto A , S e γ fornecem uma indicação da textura da galáxia. Demonstramos que os índices de tipo concentração têm uma capacidade maior para separar as galáxias *early-type* das *late-type*. Porém a correlação com a sequência de Hubble é maior para os índices de tipo textura.

No espaço dos parâmetros formado por G , γ , $\log(A)$ e M_{20} , apresentamos quatro classificações morfológicas automáticas usando os seguintes algoritmos: a regressão linear, o modelo de mistura de gaussianas, o modelo de mistura de distribuições empíricas e o método k-médias. Mostramos que a escolha do algoritmo de classificação não afeta a qualidade do resultado. Todos os métodos estimam o tipo de Hubble com um erro médio similar.

Como α mede a deformação externa da galáxia, pode ser usado para determinar um “indicador de chance de interação”. Associamos α com A , também usado na literatura para identificar interações, e definimos: $I.C.I.=A \cdot \alpha$. Aplicamos esse indicador em dois catálogos de galáxias HII. I.C.I tem uma correlação maior com γ do que com a cor ou a largura equivalente de $H\beta$. Deduzimos que a formação estelar não é máxima quando a galáxia se apresenta muito deformada; ela torna maior em uma fase posterior, quando as instabilidades dinâmicas causam um fluxo de gás significativo para a região central, como se a formação estelar respondesse à morfologia perturbada com um certo atraso.

Em um estudo espectroscópico de galáxias *late-type*, comparamos onze métodos diferentes de determinação da metalicidade do gás por linhas de emissão. Pelo método de síntese de população, calculamos as massas, idades e metalicidades das populações parciais. Observamos que uma primeira amostra, formada de galáxias com alta taxa de formação estelar, segue a relação massa-metalicidade, ao contrário de uma segunda amostra, formada de galáxias HII selecionadas por critérios muito mais estritos. O enriquecimento da metalicidade das galáxias HII mostra uma ligeira dependência com a massa, quando maior do que $5 \cdot 10^9 M_{\odot}$. Estas observações favorecem modelos de evolução galáctica com constante queda de gás (*gas infall*).

Abstract

The main purpose of this thesis is to deal with the surface photometry of large data samples. We have implemented our own fully automated processes of data reduction and the computation of photometric parameters. A large part of this work consists in the technical description of the developed methods.

We studied the morphology using a catalog of 2253 galaxies, whose Hubble type is determined. The photometric quantities are computed within the “Petrosian isophote”, a photometric aperture combining the advantages of an isophote and a characteristic radius. We calculated the usual morphological indices: the Sérsic parameter n , the indices of concentration C , asymmetry A and clumpiness S , the Gini coefficient G and the M_{20} index. We tried to improve the computing processes, for example using systematically elliptical apertures.

We define the anellipticity index α as the evaluation of the outer galaxy isophote deformation, and the granulation index γ as the slope of the spectrum formed by the second-order radial moments of the wavelet coefficients; γ can be interpreted as the feature of the galaxy to contain star-forming regions far from its center.

The correlations between indices lead to the conclusion for the existence of two parametric families: n , C and G as a light concentration measure; and A , S and γ as a galaxy texture indication. We show that concentration-like indices have a better ability to separate early- from late-type galaxies; while texture-like indices correlate best with the Hubble sequence.

In the parameter space formed by G , γ , $\log(A)$ and M_{20} , we present four automatic morphological classifications using the following algorithms:

the linear regression, the Gaussian mixture model, the empirical distributions mixture model, and the k-means method. We show that the choice of the classification algorithm does not affect the quality of the result. All methods estimate the Hubble type with a similar mean error.

As α measures the outer galaxy deformation, it can be used to determine an “interaction chance indicator”. We associate γ with A , also used in the literature to identify interactions, and define: $I.C.I=A \cdot \alpha$. We apply this indicator onto two HII galaxy catalogues. I.C.I has a higher correlation with γ as with the colour or the equivalent width of $H\beta$. We deduce that star formation is not maximum when the galaxy is very deformed. The star formation becomes larger in a later stage, when dynamic instabilities cause a significant gas infall to the central region, just as if star formation responded to disturbed morphology with a delay.

In a spectroscopic study of late-type galaxies, we compared eleven different methods of line-based gas metallicity determination. By population synthesis, we calculate the partial populations masses, ages and metallicities. We observed that a first sample of high star formation rate galaxies follows the mass-metallicity relation, unlike a second sample, formed by HII galaxies selected by more stringent criteria. The HII galaxies metallicity enrichment shows a slight dependence on the mass, when greater than $5 \cdot 10^9 M_{\odot}$. These observations favour galaxy evolution models featuring constant gas infall.

Conteúdo

Lista de Figuras	xiii
Lista de Tabelas	xix
1 Introdução	1
1.1 Visão histórica	1
1.2 Morfologia e interação	7
1.3 Motivação e objetivos	11
2 Dados e processamento preliminar de imagem	13
2.1 Catálogos	13
2.1.1 O <i>Sloan Digital Sky Survey</i>	14
2.1.2 O catálogo morfológico de Fukugita	15
2.1.3 Os catálogos de galáxias HII	18
2.2 Redução dos dados fotométricos	19
2.2.1 Os objetos fotométricos do SDSS	19
2.2.2 Limpeza das imagens	25
2.2.3 Subtração do céu	33
2.3 Abertura	35
2.3.1 Aproximação com um <i>threshold</i>	36
2.3.2 Os momentos de uma imagem	37
2.3.3 O raio de Kron	40
2.3.4 O raio de Petrosian	40
2.3.5 A isofota de Petrosian	41

CONTEÚDO

3	Morfologia quantitativa	45
3.1	Correlações entre os índices	47
3.2	Índices comuns	48
3.2.1	O parâmetro de Sérsic	48
3.2.2	O índice de concentração	52
3.2.3	O índice de assimetria	55
3.2.4	O índice de <i>clumpiness</i>	59
3.2.5	O coeficiente de Gini	62
3.2.6	O índice M_{20}	66
3.3	Dois novos índices morfológicos	70
3.3.1	O índice de anelipticidade	70
3.3.2	O índice de granulação	73
4	Classificações morfológicas automáticas	83
4.1	Introdução	83
4.2	As nossas classificações	85
4.2.1	Regressão linear	88
4.2.2	Distribuição normal multivariada	91
4.2.3	Distribuição empírica	96
4.2.4	Método k-médias	100
5	Anellipticity and granulation as morphological indices to explore star formation in interacting galaxies.	103
6	Gas and stellar metallicities in HII galaxies	125
7	Conclusão	137
A	Representações gráficas e pontos divergentes	141
B	Medindo as desigualdades	147
C	Wavelet vs Fourier	157
	Referências	163

Lista de Figuras

1.1	Galáxias ou não? Primeira tentativa de classificação das “nebulosas” por Wolf (1908).	3
1.2	A sequência de Hubble em diapasão.	4
1.3	A descrição morfológica tridimensional de de Vaucouleurs (1959). . .	5
1.4	Vertical corte da imagem precedente, como foi publicado em de Vaucouleurs et al. (1964).	6
1.5	Exemplos de galáxias em interação. Crédito: <i>The Hubble Heritage Project</i>	8
1.6	Acima: galáxia dos Ratos (Crédito: Adam Block/Mount Lemmon SkyCenter/University of Arizona). Baixo: simulação de Toomre & Toomre (1972).	9
1.7	<i>NGC3314</i> não é um par de galáxias em interação, mas duas galáxias na mesma linha de visada. <i>NGC 3314a (face-on)</i> fica a uma distância de cerca 35 Mpc, à frente de <i>NGC 3314b</i> localizada a 42 Mpc. Crédito: <i>The Hubble Heritage Project</i>	10
2.1	Esquerda: O telescópio do SDSS (Crédito: David Kirkby). Direita: As 30 câmeras CCD do SDSS. Cada uma das seis linhas fotografa uma listra no céu nas cinco bandas, da esquerda para direita: <i>g, z, u, i</i> e <i>r</i> (Crédito: SDSS).	14
2.2	Histogramas das magnitudes na banda <i>r</i> (esquerda) e dos <i>redshifts</i> (direita) para o catálogo morfológico de Fukugita.	16
2.3	Histograma dos tipos morfológicos <i>T</i> para o catálogo morfológico de Fukugita.	17

LISTA DE FIGURAS

2.4	<i>Fukugita0048</i> e <i>Fukugita0049</i> como aparecem nos arquivos <i>fpAtlas</i> do SDSS.	20
2.5	A imagem original da galáxia (esquerda) é separada em quatro objetos pelo <i>deblender</i> do SDSS. A classificação automática do SDSS identifica os dois primeiros como estrelas e os dois últimos como galáxias. .	21
2.6	A região da galáxia considerada, conforme SDSS. Todos os objetos fotométricos detectados estão circulados em azul. A linha verde contém todos os objetos <i>children</i> de um mesmo <i>parent</i>	22
2.7	Imagem do campo <i>fpC-000752-g4-0284</i> . A galáxia <i>Fukugita1130</i> é extraída cortando um quadrado com $6r_{Petro}$ de lado.	24
2.8	A galáxia <i>Fukugita1130</i> na banda <i>g</i>	25
2.9	Histograma da Imagem 2.8 (linha preta) e o ajuste por uma lei de Poisson (linha vermelha). O máximo da curva determina o <i>threshold</i> . O histograma preenchido em cinza é o do céu, normalizado com o mesmo número de píxeis.	27
2.10	Na esquerda, a imagem binária da Figura 2.8: píxeis maiores que o <i>threshold</i> estão em branco, menores em preto. No meio, a mesma imagem após a erosão e a dilatação. Na direita, a máscara conserva apenas o objeto central.	28
2.11	Função de Poisson com o parâmetro livre que define o valor do <i>threshold</i>	29
2.12	A galáxia <i>Fukugita1130</i> com os contornos das três máscaras resultantes dos três <i>thresholds</i> da Figura 2.11; o mais baixo (-8.44) é o mais externo.	29
2.13	Procedimento completo de limpeza das imagens, exemplo com <i>Fukugita1130</i> . Na última linha, a imagem limpa e todos os objetos removidos.	31
2.14	Mapa de segmentação computado pelo SDSS (esquerda) e aquele que decorre do nosso procedimento de limpeza (direita).	32
2.15	Imagem em <i>r-i</i> , <i>r-z</i> , <i>i-z</i> da galáxia <i>Fukugita1130</i>	32

2.16 Corte ao longo do eixo X da galáxia <i>Fukugita1130</i> . A linha cinza corresponde aos píxeis da galáxia, enquanto a preta corresponde ao céu. As retas verticais são os limites da máscara da galáxia após (linha contínua) e antes (linha tracejada) a dilatação. A linha vermelha é o nível do céu, resultante de um ajuste linear da curva preta. A inclinação do céu é exagerada para maior clareza.	34
2.17 Corte da galáxia <i>Fukugita1130</i> ao longo do eixo X (linha preta). Calcule-se os momentos de 1 ^a e 2 ^a ordens para definir a gaussiana (linha vermelha). A borda a 3σ da gaussiana define uma elipse que pode ser usada como abertura da galáxia.	39
2.18 Função de Butterworth usada como filtro passa-baixo na imagem antes de computar a isofota de Petrosian.	42
2.19 Todas as aberturas desenvolvidas em nosso trabalho. Usamos a isofota de Petrosian para as medidas fotométricas.	44
3.1 Processo de computação do índice de Sérsic n ; três exemplos de uma galáxia elíptica (<i>Fukugita1849</i>), uma espiral (<i>Fukugita0443</i>) e uma irregular (<i>Fukugita0022</i>).	50
3.2 Esquerda: distribuição de n . A linha preta se refere ao catálogo morfológico inteiro, o histograma vermelho às galáxias <i>early-type</i> e o azul às galáxias <i>late-type</i> . Direita: valor de n em função do tipo morfológico. Os quadrados representam a mediana para cada tipo e as barras indicam a dispersão de um desvio absoluto mediano.	51
3.3 Processo de computação do índice de concentração C ; três exemplos sobre <i>Fukugita1849</i> , <i>Fukugita0443</i> e <i>Fukugita0022</i>	53
3.4 Distribuição de C (esquerda) e mediana de C em função do tipo morfológico (direita).	54
3.5 Relação entre C e n . A cor dos pontos indica o tipo morfológico tal como determinado por Fukugita et al. (2007).	55
3.6 Processo de computação do índice de assimetria A	56
3.7 Distribuição de A (esquerda) e mediana de A em função do tipo morfológico (direita).	57
3.8 Relações entre A e os índices precedentes.	58

LISTA DE FIGURAS

3.9	Processo de computação do índice de clumpiness S	60
3.10	Distribuição de S (esquerda) e mediana de S em função do tipo morfológico (direita).	61
3.11	Relações entre S e os índices precedentes.	61
3.12	Processo de computação do coeficiente de Gini G	63
3.13	Distribuição de G (esquerda) e mediana de G em função do tipo morfológico (direita).	64
3.14	Relações entre G e os índices precedentes.	65
3.15	Processo de computação do índice M_{20}	67
3.16	Distribuição de M_{20} (esquerda) e mediana de M_{20} em função do tipo morfológico (direita).	68
3.17	Relações entre M_{20} e os índices precedentes.	69
3.18	Processo de computação do índice de anelipticidade α	72
3.19	Distribuição de α (esquerda) e mediana de α em função do tipo morfológico (direita).	73
3.20	Relações entre α e os índices precedentes.	74
3.21	Esquerda: Decomposição em <i>wavelets</i> da galáxia <i>Fukugita0443</i> . Direita: Os coeficientes de <i>wavelet</i> são multiplicados por r^2 para obter o momento central de segunda ordem. Os fluxos totais em cada uma destas imagens formam o espectro de potência.	76
3.22	Processo de computação do índice de granulação γ	77
3.23	Distribuição de γ (esquerda) e mediana de γ em função do tipo morfológico (direita).	78
3.24	Relações entre γ e os índices precedentes.	80
4.1	Comparação da classificação visual efetuada por Nair (notada de PN) com aquelas de RC3 (esquerda) e de Fukugita (direita). Figura tirada de Nair & Abraham (2010).	86
4.2	Esquerda: Comparação entre o tipo morfológico determinado por uma regressão linear e o tipo de referência. Direita: Uma segunda regressão, inversa, ajusta novamente o tipo morfológico. A posição dos pontos em X é ligeiramente deslocada para mais clareza. A linha na diagonal indica a igualdade.	89

4.3	Comparação do tipo determinado por regressão e a classificação de Fukugita para a amostra apurada (esquerda) e a amostra de trabalho (direita). O tamanho dos círculos é proporcional ao número de galáxias.	89
4.4	Comparação entre classificação pelo método da distribuição normal multivariada e o tipo de Fukugita para a amostra apurada (esquerda) e a amostra de trabalho (direita).	96
4.5	Mapas de probabilidade, exemplo no plano $G-\gamma$.	97
4.6	Comparação entre o método da distribuição empírica e o tipo de Fukugita para a amostra apurada (esquerda) e a amostra de trabalho (direita).	98
4.7	Comparação entre a classificação pelo método das k-médias e o tipo de Fukugita para a amostra apurada (esquerda) e a amostra de trabalho (direita).	102
A.1	A três variáveis consideradas ($u-i$, γ e o tipo morfológico T), uma em função da outra. A variável T é artificialmente deslocada para mais clareza.	142
A.2	Nuvem de pontos no plano cor-granulação. A cor dos pontos indica o tipo morfológico.	142
A.3	Plano cor-granulação para os tipos morfológicos. À esquerda: média com barra de 1 desvio padrão; à direita, mediana com barra de 1 desvio absoluto médio.	143
A.4	Plano cor-granulação para os tipos morfológicos. À esquerda: a localização e a dispersão dos tipos morfológicos são representados por elipses cujo eixos são as PCs, com 1σ de comprimento. À direita: os PCs são calculados com a pseudo-covariância, ou seja, com medianas e desvios absolutos medianos.	145
A.5	O plano cor-granulação expurgado dos pontos divergentes a partir da utilização do método PCA, usando pseudo-covariância (esquerdo). Sem pontos divergentes, as elipses podem ser calculadas de maneira normal, com médias e desvios padrões (direita).	146

LISTA DE FIGURAS

B.1	Esquerda: Curva de Lorenz de renda familiar no Brasil, em 2002 (Fonte: Pesquisa de orçamentos familiares, IBGE). Direita: Evolução do coeficiente de Gini do Brasil durante os últimos anos (Fonte: Pesquisa nacional por amostra de domicílios, IBGE).	148
B.2	Mapa do mundo com a cor dos país indicando o coeficiente de Gini.	149
B.3	Curvas de Lorenz para os exemplos de altura (linha azul) e de rendimento (linha vermelha). A linha tracejada é a linha de perfeita igualdade. As áreas cinzas definem os coeficientes de Gini.	151
B.4	Três “galáxias” com coeficientes de Gini basicamente idênticos. Esquerda: a galáxia NGC 2715. Meio: Imagem obtida por justaposição da galáxia original com uma mesma galáxia ligeiramente virada e deslocada. Direita: Imagem construída permutando aleatoriamente os pixels da galáxia. Figuras tiradas de Abraham et al. (2003).	153
B.5	O coeficiente de Gini G é muito bem aproximado pelo quociente σ_I/\bar{I}	154
C.1	Dois exemplos de <i>wavelet</i>	158
C.2	Dois bancos de filtros. O primeiro é composto de funções retangulares, o segundo de diferencias de duas tangentes hiperbólicas.	159
C.3	Três exemplos de decomposição da imagem <i>Fukugita1130</i> : com a transformada de Fourier e o banco de filtros retangulares (1ª coluna), com a transformada de Fourier e o banco de filtros tangentes hiperbólicas (2ª coluna), e com a transformada de <i>wavelet</i> (3ª coluna).	160

Lista de Tabelas

2.1	Bandas do SDSS.	15
2.2	Tipos morfológicos no catálogo de Fukugita.	16
4.1	Tipos morfológicos pelo método da dupla regressão linear.	90
4.2	Tipos morfológicos pelo método da distribuição normal multivariada.	95
4.3	Tipos morfológicos pelo método de uma distribuição empírica.	99
4.4	Tipos morfológicos pelas k-médias.	101
B.1	Exemplo de 10 indivíduos cujas alturas e salários são conhecidos.	150
B.2	Somas acumuladas e normalizadas das sequências ordenadas.	151

LISTA DE TABELAS

Capítulo 1

Introdução

1.1 Visão histórica

O poeta e naturalista alemão Johann Wolfgang von Goethe (1749 – 1832) interessou-se em plantas e tentou descrevê-las de maneira científica. No entanto, ele não ficou satisfeito com o método comum de descrição como uma soma de vários elementos e procurou estabelecer uma representação mais global. Confrontado com a dificuldade de qualificar essa “maneira diferente” de apreender o mundo vegetal, ele construiu uma nova palavra unindo do grego μορφή (morphé), que significa “forma”, com o sufixo padrão λόγος (lógos), que significa “ciência”. A morfologia é, então, a ciência da forma.

Em 1817, Goethe explicou mais detalhadamente em *Zur Naturwissenschaft überhaupt, besonders zur Morphologie* o conceito a partir do qual a nova disciplina foi criada:

« Quando nós nos apercebemos dos objetos naturais e sobretudo dos objetos vivos, de modo que desejamos proporcionar uma compreensão do conjunto do seu ser e da sua atividade, cremos chegar da melhor maneira a tal conhecimento através da dissociação das partes; [...] Constantemente efetuados, porém, estes esforços de dissociação produzem também muitos inconvenientes. O ser vivo pode ser decomposto nos seus elementos, mas a partir deles não se pode reconstituí-lo e devolver-lhe a vida. [...] É por isso que, em todas as épocas, também se manifestou no homem de ciência um impulso para reconhecer as formações vivas enquanto tais, de

1. INTRODUÇÃO

apreender as suas partes exteriores tangíveis e visíveis, para as aceitar como indícios e, assim, dominar de certo modo o todo na intuição. [...] Encontramos, por conseguinte, no curso da arte, do saber e da ciência, várias tentativas para fundar e desenvolver uma doutrina, a que gostaríamos de chamar Morfologia. »¹

Com certeza, essa definição da morfologia fica bastante abstrata e quase filosófica. Mas o ponto central está claramente expresso: as partes tangíveis e visíveis estão a serviço de uma compreensão do conjunto. Essa metodologia consiste numa abordagem qualitativa baseada na fenomenologia.

A descrição é a primeira etapa para o entendimento de um objeto, sua formação e sua evolução. Com esse propósito, os astrônomos não esperaram saber o que eram os objetos difusos e nebulosos do céu para poder descrevê-los. Já no início do século XX, Wolf (1908), num artigo cujo título é “Die Klassifizierung der kleinen Nebelflecken” (A classificação das pequenas nebulosas, “Flecken” = “manchas”, “Nebel” = “nevoeiro” ou “nebulosa”), descreve e tenta classificar o que se chamava na época de nebulosas. A Figura 1.1 representa os seus desenhos originais.

Em 1908, o “grande debate” não fora fechado; na verdade, ainda não tinha sido aberto. Naquela época, não se conhecia a natureza física dos objetos difusos, e, portanto, a classificação de Wolf não fazia distinção entre objetos astronômicos diferentes: galáxias, aglomerados estelares e nebulosas planetárias. Todos esses objetos estão misturados naquela classificação.

Somente no meio dos anos 1920 Edwin Hubble encerra o “grande debate”, medindo os períodos das estrelas variáveis cefeidas na galáxia de Andrômeda, o que permitiu calcular precisamente a sua distância. Doravante, não havia mais dúvidas: as “nebulosas espirais” eram galáxias semelhantes à Via Láctea. A partir de então, uma classificação de galáxias propriamente dita pôde ser desenvolvida. Veja Sandage (1975) para uma ampla história das classificações morfológicas de galáxias.

¹Tradução de Maria Molder em “Goethe, J. W. *A metamorfose das plantas*”. Lisboa: Imprensa Nacional - Casa da Moeda, 1993.

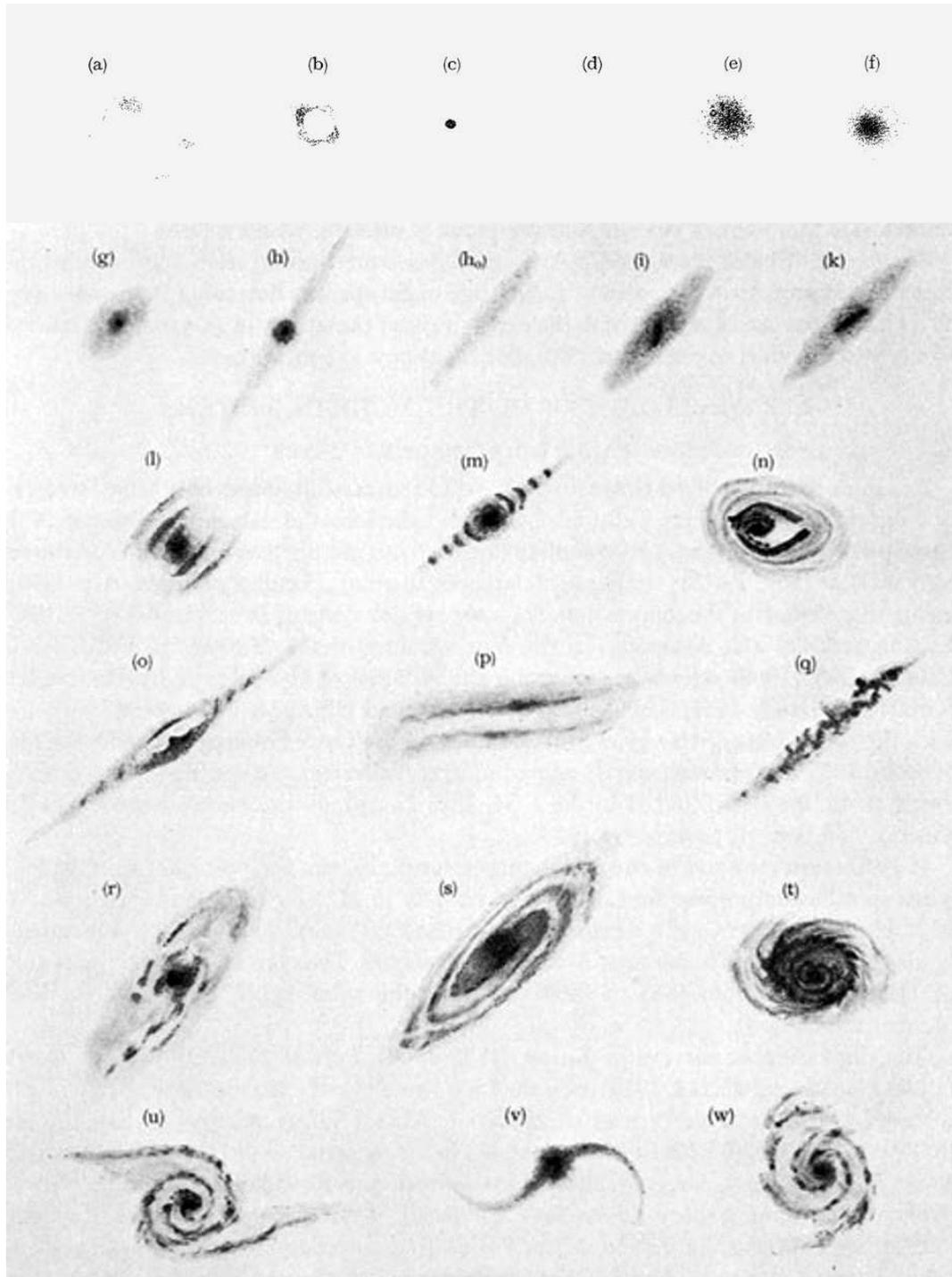


Figura 1.1: Galáxias ou não? Primeira tentativa de classificação das “nebulosas” por Wolf (1908).

1. INTRODUÇÃO

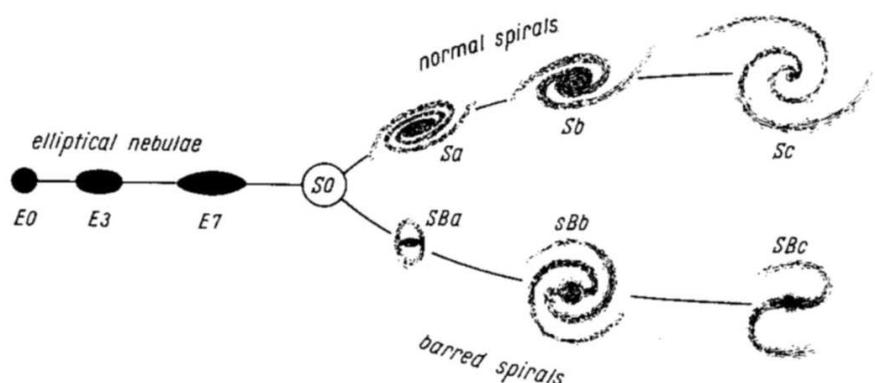


Figura 1.2: A sequência de Hubble em diapásio.

Hubble (1926) propõe a sua célebre sequência de “diapasão” (*tuning fork*, em inglês) em um artigo chamado *extra-galactic nebulae*. A ilustração publicada no seu livro, dez anos depois, (Hubble 1936) é reproduzida na Figura 1.2.

A sequência de Hubble especifica os diferentes tipos:

- As galáxias elípticas, denotadas por “E”, seguido de um parâmetro definindo a elipticidade;
- As galáxias lenticulares, denotadas por “S0”;
- Separação em espirais normais, denotadas por “S”, e espirais barradas, denotadas por “SB”;
- As galáxias espirais “Sa” (SBa), com braços espirais difusos e muito enrolados, e um bojo grande e luminoso;
- As galáxias espirais “Sb” (SBb), com braços menos enrolados e um bojo pequeno e fraco;
- As galáxias espirais “Sc” (SBc), com braços pouco enrolados e fragmentados, permitindo discernir regiões de formação estelar, e um bojo ainda menor e mais fraco.

A base do sucesso da classificação de Hubble é o fato que ela se organiza em uma sequência, ou seja, uma lista ordenada. Embora subjetiva, uma quantificação intuitiva pode facilmente ser adotada, usando um único parâmetro T . Na ideia de uma sequência organizada, as galáxias elípticas e lenticulares são chamadas de *early-type* (tipo inicial) e as espirais, de *late-type* (tipo tardio). Essas denominações foram impropriamente atribuídas, pensando-se que a sequência de Hubble era evolutiva.

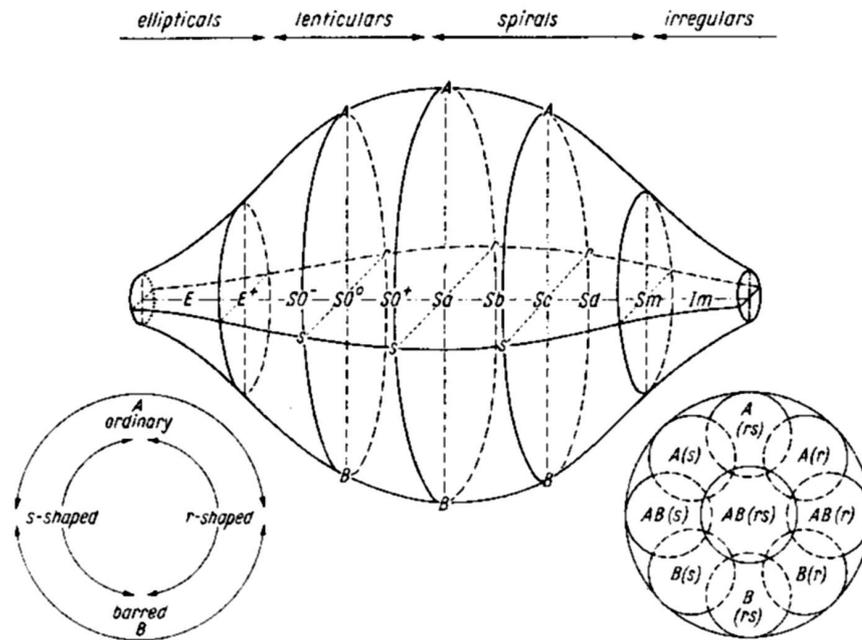


Figura 1.3: A descrição morfológica tridimensional de de Vaucouleurs (1959).

Mais tarde, na direita da sequência de Hubble, dois novos tipos foram adicionados por de Vaucouleurs (1959):

- As galáxias “Sd” (SBd) ou “Sm”, com braços espirais muito fragmentados e um bojo fraco ou inexistente;
- As galáxias “Im”, para “irregulares tipo Magalhães”, que não possuem estruturas espirais, mas cujas regiões de formação estelar parecem ser espalhadas aleatoriamente no disco.

De Vaucouleurs introduziu mais uma extensão à classificação acima, com o objetivo de indicar a presença de um anel interno. A sua descrição morfológica se compõe, então, de três parâmetros, o primeiro descrevendo os braços espirais, o segundo para a barra e o terceiro para o anel. A morfologia tridimensional de de Vaucouleurs é representada na Figuras 1.3 e 1.4; imagens publicadas em, respectivamente, de Vaucouleurs (1959) e de Vaucouleurs et al. (1964).

Outras descrições morfológicas foram adaptadas, por exemplo por Morgan (1958), van den Bergh (1960a,b) ou Hodge (1966). Em seguida, apareceram catálogos de

1. INTRODUÇÃO

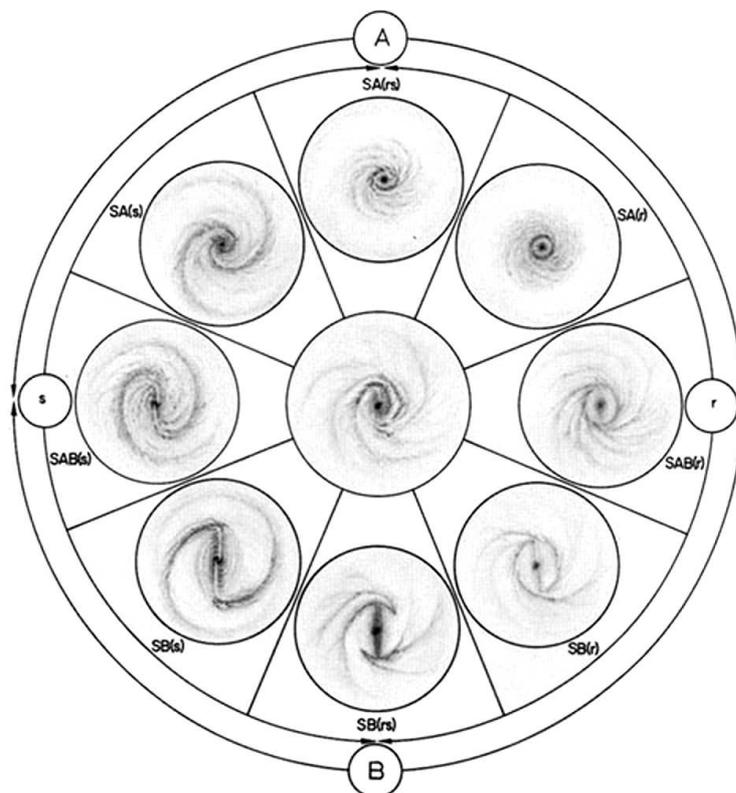


Figura 1.4: Vertical corte da imagem precedente, como foi publicado em de Vaucouleurs et al. (1964).

galáxias cujo tipo morfológico foi visualmente determinado. Os exemplos mais conhecidos são o *Reference Catalogue of Bright Galaxies*, primeira, segunda e terceira versões (de Vaucouleurs et al. 1991, 1976, 1964), o *Uppsala General Catalogue of Galaxies* (Nilson 1973) ou o *Revised Shapley Ames Catalogue* (Sandage & Tammann 1980). Outros catálogos são especialmente dedicados a uma certa classe de galáxias, como o *Atlas of Peculiar Galaxies* (Arp 1966) ou o *Atlas and catalog of interacting galaxies* Vorontsov-Velyaminov (1959) para as galáxias irregulares ou em interação.

Hoje, dois pontos de vista – quase duas filosofias – se enfrentam na comunidade científica, de acordo com as classificações morfológicas: os adeptos das classificações visuais e aqueles das classificações automatizadas. As classificações automáticas se tornaram possíveis com introdução das imagens CCD e a generalização do tratamento computacional. Utiliza-se, às vezes, o termo de morfometria para definir os métodos

de medida da morfologia. Mesmo em trabalhos recentes, pesquisadores preferem o método tradicional e fastioso da inspeção visual. A argumentação dada por Nair & Abraham (2010), que classificaram visualmente mais de 14 000 galáxias, é representativa de uma certa opinião:

« Considerable progress has been made in developing techniques for automated classification of galaxies. However, the disappointing fact is that, at present, state-of-the-art automated galaxy classification is only capable of delivering crude classifications, albeit very quickly. The large numbers of classifications delivered by automated techniques have proved highly useful in their appropriate context, but they are neither as accurate, nor as comprehensive, as visual classifications made by a trained observer. The strengths and limitations of present machine-based galaxy classification techniques should not come as a surprise if one reflects upon the fact that something like half the human brain is devoted to vision. Millions of years of evolution have refined our brain's capacity for image processing and analysis to the extent that a well-trained human can classify an individual image more accurately than a computer-based algorithm in essentially every field of science or industry in which comparisons have been made. »

A origem da sequência de Hubble foi por muito tempo mal compreendida. Há uma década, estabeleceu-se um consenso científico ao redor do papel central das fusões na evolução galáctica. Os tipos morfológicos hoje observados são resultados da história de fusões passadas entre galáxias. Cada configuração de encontro, notavelmente as fusões grandes (*major mergers*), transformam profundamente a morfologia. Veja Benson (2010) ou Buta (2011) para recentes revisões sobre a formação e a evolução galáctica.

1.2 Morfologia e interação

Galáxias interativas são coisas correntemente observadas na história cósmica. A presença de uma massa vizinha causa perturbações gravitacionais, que modificam significativamente a morfologia da galáxia. As interações desempenham um papel capital na evolução galáctica e uma das aplicações mais comuns dos estudos morfológicos consiste na sua detecção.

1. INTRODUÇÃO

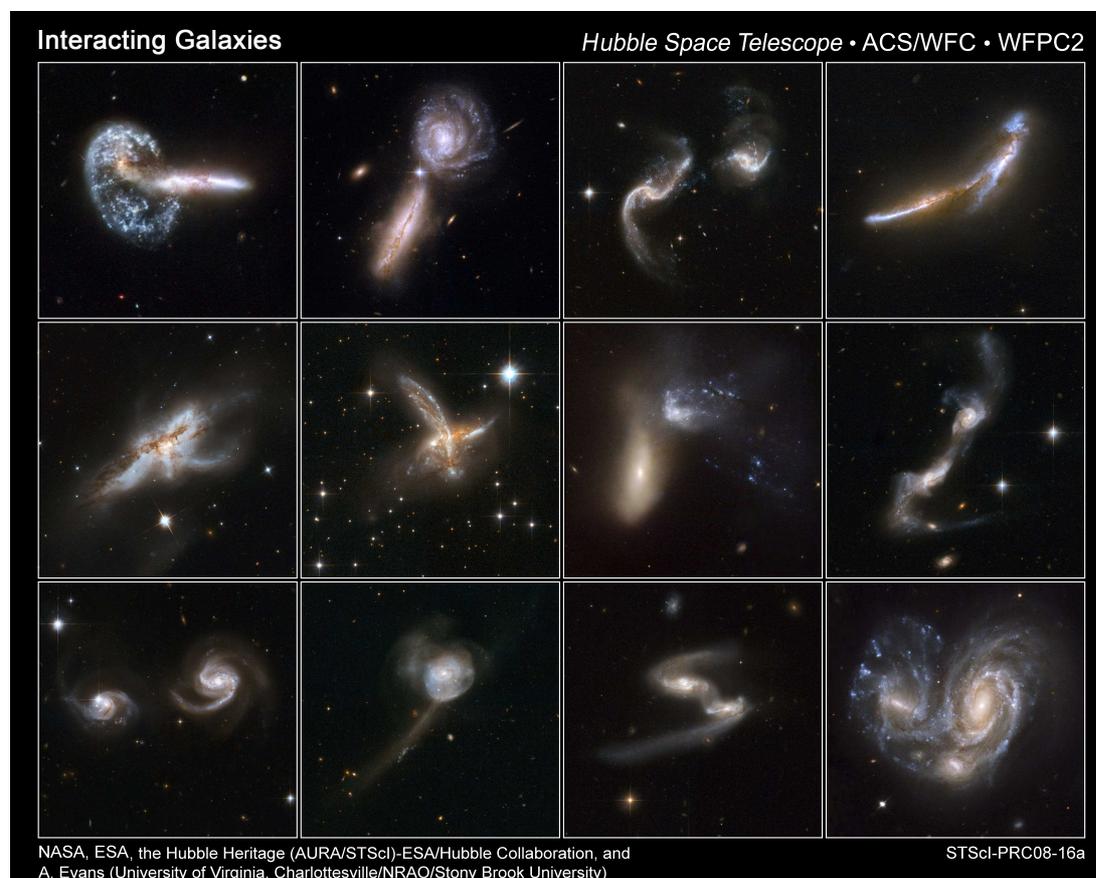


Figura 1.5: Exemplos de galáxias em interação. Crédito: *The Hubble Heritage Project*.

Interação é um termo geral para denominar o encontro de duas – ou mais! – galáxias. Fala-se de fusão quando são envolvidas condições para a formação de um única galáxia no final. A palavra colisão é, muitas vezes, mas não sempre, reservada para situações nas quais cada uma das galáxias prossegue o seu caminho no espaço.

A Figura 1.5 mostra exemplos de galáxias em interação do *Hubble Heritage Project* (<http://heritage.stsci.edu>). As galáxias são: *Arp148*, *UGC9618*, *Arp256*, *NGC6670*, *NGC6240*, *ESO593-8*, *NGC454*, *UGC8335*, *NGC6786*, *NGC17*, *ESO77-14* e *NGC6050*.

Durante a interação, as estrelas não entram em colisão, mas as órbitas são deformadas por efeito de maré, gerando a criação de pontes e caudas. Essas estruturas podem ser detectadas através de medidas morfológicas, que atuam como uma assinatura da interação.

Ao contrário das estrelas, o gás é colisional. As perturbações gravitacionais in-

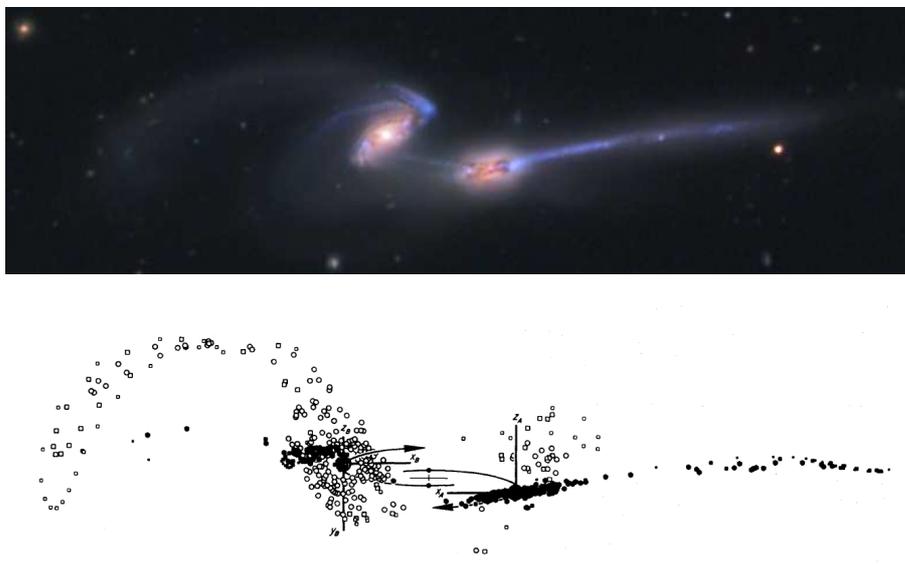


Figura 1.6: Acima: galáxia dos Ratos (Credito: Adam Block/Mount Lemmon SkyCenter/University of Arizona). Baixo: simulação de Toomre & Toomre (1972).

troduzem fricção no meio interestelar, que se comprime e se aquece. As nuvens de gás molecular colapsam, e levam à produção de uma nova geração de estrelas. A formação estelar é, assim, desencadeada pela fusão ou interação (*interaction- or merger-triggered star formation*).

Uma grande parte da nossa compreensão das interações galácticas provém das simulações. O trabalho precursor de Toomre & Toomre (1972), embora sendo por parâmetros atuais uma simples simulação de N-corpos, forneceu resultados importantes. No exemplo da Figura 1.6, eles tentam modelar a galáxia *NGC4676*, chamada de galáxia dos Ratos. As simulações mais recentes incluem todos os componentes de uma galáxia (estrelas, gás H_I , gás H_{II} , gás molecular, poeira, matéria escura) em modelos hidrodinâmicos de partículas suavizadas (*smoothed particle hydrodynamics*), assumindo a lei de Schmidt-Kennicutt (Kennicutt 1989) para a formação estelar. Podemos citar, como exemplo, a série de artigos de Lotz et al. (2008, 2010a,b) que estuda a evolução da morfologia durante uma fusão, calculando índices morfológicos nas imagens simuladas.

Galáxias em interação podem ser detectadas por distúrbios na sua morfologia. Contudo, temos que ter cuidado com o fato que a fotometria é dependente da geo-

1. INTRODUÇÃO

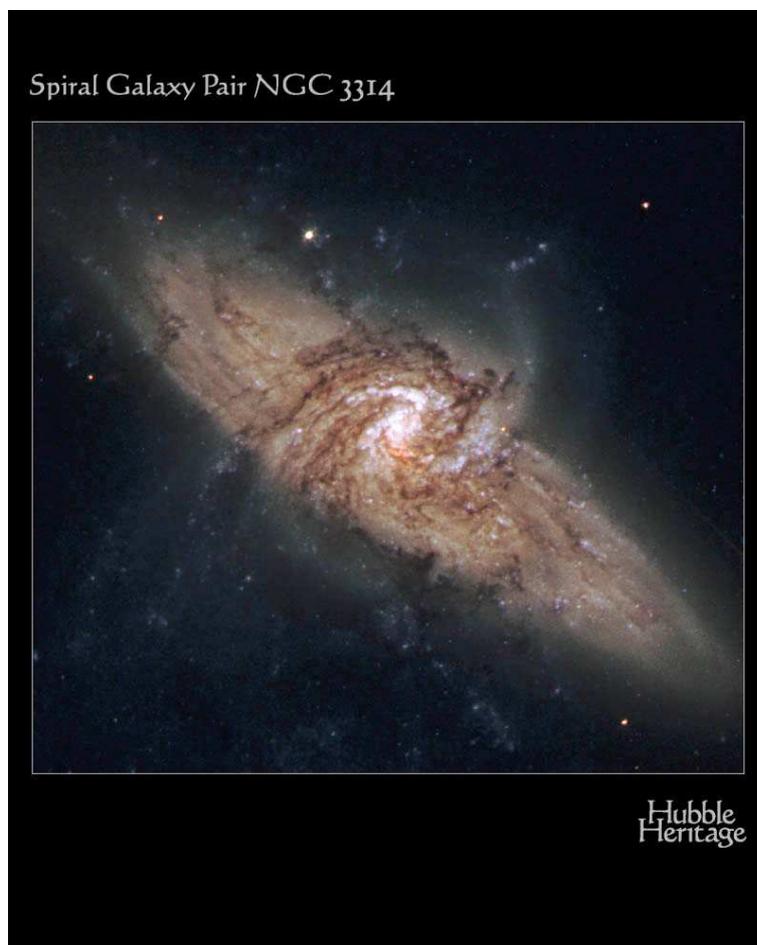


Figura 1.7: *NGC3314* não é um par de galáxias em interação, mas duas galáxias na mesma linha de visada. *NGC 3314a (face-on)* fica a uma distância de cerca 35 Mpc, à frente de *NGC 3314b* localizada a 42 Mpc. Crédito: *The Hubble Heritage Project*.

metria do sistema, pois temos acesso apenas a uma projeção na esfera celeste. As aparências podem enganar: duas galáxias aparentemente vizinhas podem estar, na realidade, muito afastadas uma da outra. A Figura 1.7 ilustra essa problemática com *NGC3314*. Esse sistema é formado de duas galáxias espirais separadas de 5 a 8 Mpc, mas localizadas exatamente na mesma linha de visada. Para os observadores terrestres que somos, se assemelham, à primeira vista, a galáxias em interação.

1.3 Motivação e objetivos

Neste trabalho, propomo-nos a conduzir um estudo completo da morfologia quantitativa das galáxias. Examinamos os diferentes métodos de medida do tipo de Hubble, calculando os índices morfológicos encontrados na literatura: o parâmetro de Sérsic, os índices de concentração, de assimetria e de *clumpiness*, o coeficiente de Gini e o índice M_{20} . Analisamos os diferentes processos de cálculo e a estabilidade estatística dos índices. Definimos também dois novos índices morfológicos especialmente dedicados a galáxias com alta taxa de formação estelar, a fim de estudar, em especial, as galáxias em interação. O primeiro é o índice de anelipticidade que avalia a deformação da isofota externa de uma galáxia; o segundo é o índice de granulação que mede a distribuição das regiões de formação estelar.

Pretendemos avaliar as correlações dos índices entre si e com a sequência de Hubble, assim como a capacidade dos índices em separar as galáxias conforme o seu tipo. Com esses resultados, podemos apresentar métodos de classificações automáticas de galáxias e investigar a sua eficiência.

Definindo um indicador morfológico para detectar galáxias em interação, procuramos analisar as propriedades físicas das galáxias interativas, especialmente a localização das regiões de formação estelar, tal como avaliada pelo índice de granulação. Isso nos permite caracterizar a formação estelar durante as diferentes fases da interação.

Finalmente, estudamos a metalicidade e o enriquecimento da metalicidade para as galáxias *late-type*, para destacar o comportamento distinto de uma amostra de galáxias HII e do de uma amostra de galáxias com alta taxa de formação estelar.

1. INTRODUÇÃO

Capítulo 2

Dados e processamento preliminar de imagem

2.1 Catálogos

Com a finalidade de conduzir os diferentes trabalhos desta tese, e dependendo dos objetivos, estudamos galáxias provenientes de vários catálogos. Contudo, todas as amostras são derivadas do *Sloan Digital Sky Survey* – SDSS¹ (Stoughton et al. 2002; York et al. 2000).

Para os estudos morfológicos (Capítulos 3 e 4), usamos o catálogo de Fukugita et al. (2007), que contém 2253 galáxias, cujos tipos morfológicos foram determinados para cada uma delas visualmente pelos autores do artigo.

Em relação ao estudo sobre as galáxias em interação (Capítulo 5), desenvolvemos primeiramente o nosso método sobre esse catálogo, antes de o aplicarmos a dois catálogos de galáxias HII: O catálogo de Kniazev et al. (2004) e o catálogo de Izotov et al. (2006). Os dois juntos contêm mais de 900 galáxias, mas várias delas são de tamanho angular pequeno demais para conseguir realizar fotometria superficial.

O estudo de metalicidade das galáxias HII (Capítulo 6) é um trabalho puramente espectroscópico. Assim, não há utilização de imagens nessa parte. O catálogo de Westera et al. (2012) resulta de critérios restritos aplicados sobre as linhas de emissão.

¹<http://www.sdss.org> (site geral)
<http://cas.sdss.org> (Skyserver)

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

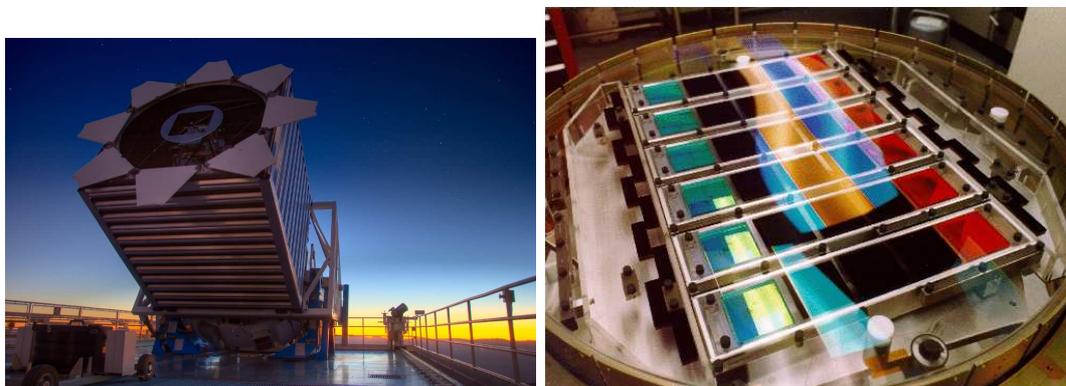


Figura 2.1: Esquerda: O telescópio do SDSS (Crédito: David Kirkby). Direita: As 30 câmeras CCD do SDSS. Cada uma das seis linhas fotografa uma listra no céu nas cinco bandas, da esquerda para direita: g , z , u , i e r (Crédito: SDSS).

Todos os espectros do SDSS foram submetidos aos critérios de seleção produzindo um catálogo final de 712 galáxias H α .

2.1.1 O Sloan Digital Sky Survey

O SDSS é um levantamento celeste cujas duas principais fases foram finalizadas em 2008 com o *Data Release 7 – DR7*² (Abazajian et al. 2009). Ele contém um catálogo de imagens de cerca de 357 milhões de objetos fotométricos. O que é chamado de objeto fotométrico não corresponde sempre a um objeto físico, tal como uma galáxia ou uma estrela, podendo causar uma certa ambiguidade (veja Seção 2.2.1).

Localizado no Observatório do Apache Point, no deserto do Novo México, nos Estados Unidos, o telescópio de 2.5 m (Figura 2.1, esquerda) cartografou o céu em um ângulo sólido de 11 663 graus quadrados, o que corresponde a quase 35% da esfera celeste. Um conjunto de 30 câmeras CCD (Figura 2.1, direita), dispostas em seis linhas, fotografa simultaneamente seis listras (*stripes*) do céu em cinco bandas, u , g , r , i , z , cujas propriedades são resumidas na Tabela 2.1. Em condições de observação

²Os novos resultados do DR8 (Aihara et al. 2011) fazem parte de uma nova fase do projeto, chamada de SDSS-III (Eisenstein et al. 2011), que tem como objetivos a busca de planetas extrassolares, o mapeamento da distribuição espacial da matéria do universo, assim como a obtenção de espectros de estrelas da Via Láctea.

Tabela 2.1: Bandas do SDSS.

Banda	Cor	$\lambda_{\text{médio}}$	m_{max}	Observação
<i>u</i>	ultravioleta próximo	3551 Å	22.0	
<i>g</i>	azul / verde	4686 Å	22.2	Contém H β , H γ e H δ para $z=0$.
<i>r</i>	vermelho	6165 Å	22.2	Contém H α para $z=0$.
<i>i</i>	infravermelho próximo	7481 Å	21.3	
<i>z</i>	infravermelho próximo	8931 Å	20.5	

normais, e com tempo de exposição de 53.9 segundos, as magnitudes atingidas m_{max} são indicadas na penúltima coluna.

As CCDs têm uma resolução angular de 0.396"/pixel. As imagens assim produzidas têm 2048 píxeis por 1489 píxeis, ou seja, 13.52'×9.83'.

O SDSS é também um levantamento espectroscópico, e o DR7 contém um total de mais de 1.6 milhão de espectros, sendo 930 000 de galáxias, 120 000 de quasares e 460 000 de estrelas. Eles cobrem os comprimentos de ondas entre 3800 e 9200 Å, com uma resolução $\lambda/\Delta\lambda$ de 1800 a 2200.

Toda noite, o SDSS produzia 200 GB de dados.

2.1.2 O catálogo morfológico de Fukugita

O catálogo de Fukugita (Fukugita et al. 2007) é composto de todas as galáxias mais luminosas do que a magnitude $m_r=16$, encontrando-se numa faixa do céu entre $145^\circ < \alpha < 236^\circ$ de ascensão reta e $-1.26^\circ < \delta < 1.26^\circ$ de declinação. Isso representa 2253 galáxias (para 2275 objetos fotométricos), com 1866 espectros registrados. Devido aos seus brilhos aparentes elevados, pode-se dizer que as galáxias ficam no universo próximo com *redshift* médio de $z_{\text{médio}}=0.056$ e máximo de $z_{\text{max}}=0.166$. As distribuições das magnitudes e dos *redshifts* do catálogo de Fukugita são representadas pelos histogramas da Figura 2.3.

Fukugita et al. (2007) seguem uma abordagem tradicional para determinar as morfologias das galáxias do seu catálogo: a classificação visual. O tipo morfológico foi determinado através de inspeção visual das imagens monocromáticas na banda *g* por cada um dos três primeiros autores do artigo (Masataka Fukugita, Osamu Nakamura

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

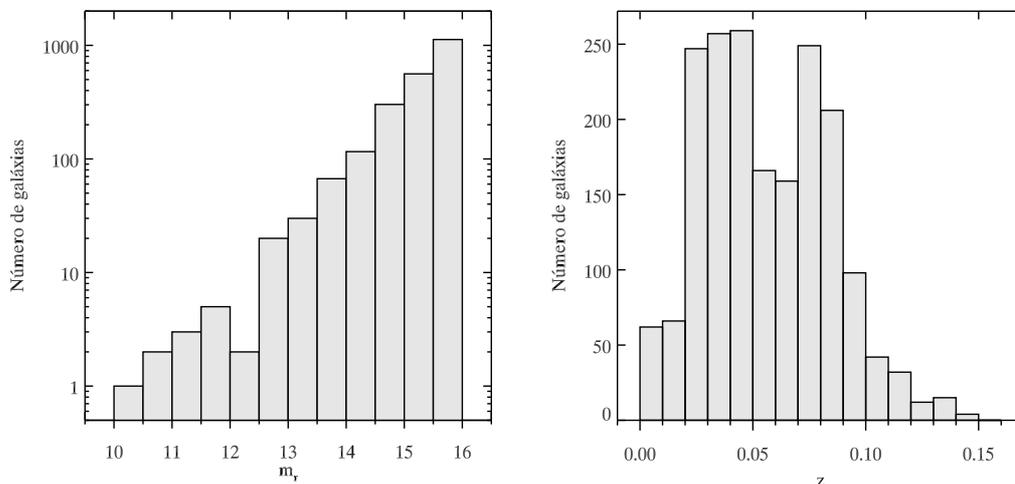


Figura 2.2: Histogramas das magnitudes na banda r (esquerda) e dos *redshifts* (direita) para o catálogo morfológico de Fukugita.

Tabela 2.2: Tipos morfológicos no catálogo de Fukugita.

Tipo de Hubble	E	S0	Sa	Sb	Sc	Sd	Im	Não classificada
T	0	1	2	3	4	5	6	-1
	0.5	1.5	2.5	3.5	4.5	5.5		

e Sadanori Okamura) independentemente. O valor final é a média redondada das inspeções para o número semi-inteiro mais próximo. Baseando-se no *Hubble Atlas of Galaxies* (Sandage 1961) ou no *Reference Catalogue of Bright Galaxies* (de Vaucouleurs et al. 1991, 1964), a classificação é definida através do tipo de Hubble por um único parâmetro T , conforme a tabela 2.2.

Não há indicação de barra. “Im” significa “Irregular, tipo Magalhães”; no *Hubble Atlas* e no *Reference Catalogue of Bright Galaxies*, essas galáxias são denominadas como “Irr I” para “Irregular, tipo I”. Outra diferença notável: o *Hubble Atlas* não inclui a classe “Sd”. O valor $T=-1$ foi atribuído para as galáxias que não se encaixam na sequência habitual de Hubble. Elas são galáxias peculiares, interatuantes ou resultantes de uma recente fusão³; no *Hubble Atlas* e no *Reference Catalogue of Bright Galaxies*,

³Os autores escrevem: “We assign $T=-1$ when we are unable to classify a galaxy into a conventional Hubble type”.

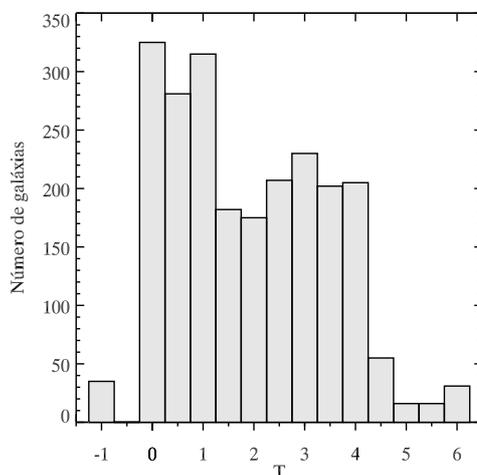


Figura 2.3: Histograma dos tipos morfológicos T para o catálogo morfológico de Fukugita.

elas são chamadas de “Irr II” e de “amorfas” em Sandage & Brucato (1979). O catálogo de Fukugita contém 33 galáxias com $T=-1$.

A distribuição do tipo morfológico é representada na Figura 2.3. Considerando as galáxias elípticas e lenticulares ($T=0-1$) como *early-type* e as galáxias espirais e irregulares tipo Magalhães ($T=1.5-6$) como *late-type*, o catálogo consiste em 921 galáxias *early-type* (41.1%) e 1319 *late-type* (58.9%). Os diferentes subtipos tradicionais de espirais (Sa, Sb e Sc) estão presentes em números equivalentes. Em contrapartida, as galáxias Sd e Im são bem menos numerosas.

Fukugita et al. (2007) identificaram 30 galáxias interagentes. Elas são frequentemente de um tipo indeterminado ($T=-1$), mas às vezes também de um tipo de Hubble. No entanto, interações gravitacionais entre galáxias não são acompanhadas sempre de assinaturas fotométricas claras, de tal maneira que o reconhecimento visual de galáxias em interação pode ser uma tarefa árdua. Certas galáxias consideradas como em interação pelos autores não parecem sê-lo para nós; e vice-versa⁴.

⁴Tentamos ilustrar essa subjetividade nas Figuras 10 e 11 do artigo do Capítulo 5.

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

2.1.3 Os catálogos de galáxias HII

Além do catálogo morfológico, catálogos de galáxias HII são usados em estudos específicos: os catálogos de Kniazev (Kniazev et al. 2004) e Izotov (Izotov et al. 2006) são utilizados na análise das interações e da formação estelar (Capítulo 5), e o catálogo de Westera et al. (2012) foi criado para o estudo das metalicidades das galáxias HII (Capítulo 6).

O catálogo de Kniazev, derivado do DR1 do SDSS, é nomeado como SHOC (*SDSS HII galaxies with Oxygen abundances Catalog*). Nele há 624 alvos espectroscópicos em 612 galáxias diferentes. Para determinar as abundâncias de oxigênio, é empregado o método padrão de temperatura eletrônica (Aller 1984).

O catálogo de Kniazev é selecionado seguindo três critérios:

- Uma de três condições das larguras equivalentes das linhas de Balmer é cumprida: $EW(H\gamma) \geq 6 \text{ \AA} \parallel EW(H\beta) \geq 20 \text{ \AA} \parallel EW(H\alpha) \geq 50 \text{ \AA}$;
- As galáxias não são AGNs (conforme o diagrama BPT);
- A abundância de oxigênio é determinada com uma precisão $\leq 0.2 \text{ dex}$.

O catálogo de Izotov contém 309 espectros do DR3 do SDSS. As abundâncias iônicas de N , O , Ne , S , Cl , Ar , assim como de Fe , foram determinadas pelo método tradicional da temperatura eletrônica e modelos recentes de fotoionização.

Os critérios de seleção para o catálogo de Izotov são:

- A linha de emissão $[OIII] \lambda 4363 \text{ \AA}$ é detectada acima do nível de 1σ do ruído;
- Os espectros de Seyfert 2 foram excluídos por inspeção visual;
- A linha $H\beta$ é forte: $F(H\beta) > 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$;
- São excluídas as galáxias com, simultaneamente,
 $[OIII] \lambda 4959/H\beta < 0.7$ e
 $[OII] \lambda 3727/H\beta > 1.0$.

O catálogo de Westera contém 712 galáxias do DR7 do SDSS. As metalicidades do gás foram determinada por 11 métodos diferentes, baseados em razões de linhas fortes como indicadores de abundância. O método que usa o parâmetro R_{23} de Pagel et al. (1979) com a calibração de Pilyugin & Thuan (2005) mostra os melhores resultados. As massas parciais e as metalicidades das populações estelares foram calculadas por síntese de população.

O catálogo de Westera é construído a partir dos seguintes critérios de seleção:

- Os espectros sem a região da linha [OII] $\lambda 3727\text{\AA}$ são descartados; isso exclui as galáxias com *redshift* menor do que 0.033;
- A razão sinal-ruído na região da linha [OII] $\lambda 3727\text{\AA}$ é alta: $S/N_{\lambda 3727} \geq 2.8$;
- As galáxias têm linhas de emissão fortes em altos níveis de excitação:
 $\log\left(\frac{[\text{OIII}] \lambda 5007 \text{\AA}}{\text{H}\beta}\right) \geq 0.2$; e, simultaneamente,
 $\log\left(\frac{[\text{NII}]}{\text{H}\alpha}\right) \leq -0.5$;
- As galáxias não são AGNs (conforme o diagrama BPT).

2.2 Redução dos dados fotométricos

Para calcular adequadamente as grandezas fotométricas, especialmente os índices morfológicos, precisamos de imagens bem preparadas. Todos os dados que usamos vêm do SDSS e são, em parte, já reduzidos. Contudo, é necessário processar as imagens das galáxias para tornar possível a fotometria superficial. Idealmente, a galáxia fica no centro da imagem, que deve ser de tamanho grande o suficiente para conter a galáxia na sua totalidade e um pouco de céu, necessário para a calibração, porém não muito grande para minimizar o tempo computacional. Para mantermos o controle durante todas as etapas do processamento, decidimos não usar SExtractor (Bertin & Arnouts 1996), IRAF (Tody 1993) ou EFIGI (Baillard et al. 2006), mas, sim, desenvolver os nossos próprios métodos. Todos os programas são escritos em IDL (*Interactive Data Language*).

2.2.1 Os objetos fotométricos do SDSS

O *Sloan Digital Sky Survey* – SDSS (Stoughton et al. 2002; York et al. 2000) é um levantamento, o que quer dizer que todas as noites o telescópio fotografa uma parte diferente da esfera celeste. O *pipeline* fotométrico (Lupton et al. 2001) analisa as imagens CCD e produz uma grande quantidade de dados. Uma parte dos algoritmos tem o papel de detectar as diferentes fontes de luz e de “desmisturar” aquelas que se sobrepõem: é o *deblender*.

Pelo *deblender* é elaborado um catálogo dos objetos fotométricos e também um atlas desses objetos para cada imagem do campo. Esse arquivo, chamado de *fpAtlas* (*fp* para *frame pipeline*), se apresenta, então, como um quebra-cabeça de pequenas

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

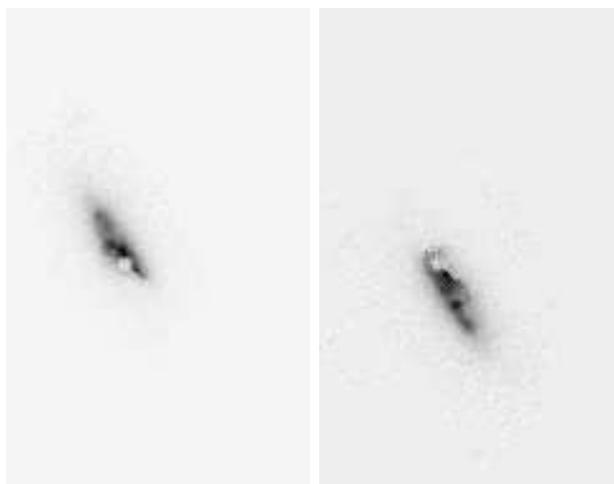


Figura 2.4: *Fukugita0048* e *Fukugita0049* como aparecem nos arquivos fpAtlas do SDSS.

imagens cuja soma é a imagem do campo original. No início do nosso trabalho, pensávamos em usar essas imagens, já separadas, sem a presença de objetos vizinhos, já que as galáxias são muito mais fáceis de manipular. No entanto logo encontramos galáxias “estranhas”.

A Figura 2.4 ilustra um exemplo de duas imagens do catálogo morfológico de Fukugita (Fukugita et al. 2007). *Fukugita0048* mostra claramente um buraco, assim como um corte súbito na sua parte inferior. *Fukugita0049* também mostra essas mesmas curiosidades, mas na parte superior. Com certeza, são artefatos computacionais.

Um controle rápido nos dados do catálogo de *Fukugita* prova que as duas galáxias estão localizadas muito próximas uma da outra, pelo menos dadas as suas ascensões retas e declinações. Apesar do *redshift* desconhecido, pois não possuem espectro, pode-se concluir que *Fukugita0048* e *Fukugita0049* são vizinhas, e até mesmo que são uma única galáxia. No entanto, os tipos morfológicos atribuídos às duas galáxias são diferentes, embora próximos. *Fukugita0048* é classificada como uma galáxia espiral Sc ($T=4$), e *Fukugita0049* como uma espiral intermediária entre Sb e Sc ($T=3.5$).

ID	SDSS-ID	RADeg	DEDeg	T	Rmag	Z
48	0756-5-44-0207-0041	146.80497	0.66636	4.0	14.99	0.0000
49	0756-5-44-0207-0040	146.80823	0.66475	3.5	15.07	0.0000



Figura 2.5: A imagem original da galáxia (esquerda) é separada em quatro objetos pelo *deblender* do SDSS. A classificação automática do SDSS identifica os dois primeiros como estrelas e os dois últimos como galáxias.

Quando buscamos todos os objetos fotométricos próximos dessas duas semigaláxias, encontramos mais duas fontes classificadas como estrelas. A Figura 2.5 mostra como a galáxia original (primeira imagem à esquerda) é separada em quatro elementos. De fato, o algoritmo do SDSS realiza uma classificação automática rudimentar, baseada no perfil de luz, para separar os objetos em estrelas e galáxias. O bojo e um aglomerado gigante são catalogados como estrelas (segunda e terceira imagens), e as partes esquerda e direita do disco realmente como galáxias (quarta e quinta imagens).

Na verdade, essa separação excessiva dos objetos é desejada, pois o *fpAtlas* é só um produto intermediário do processamento. E, portanto, o atlas não pode e não deve ser usado. Sucintamente, veremos como a *pipeline* fotométrico do SDSS trabalha.

- Os objetos muito luminosos são identificados por uma primeira detecção a $m_r < 17.5$ (cerca 200σ acima do ruído do céu).
- A fim de reduzir o ruído, a imagem é suavizada.
- Durante uma segunda detecção, todos os objetos acima de 5σ em uma das cinco bandas são catalogados no *fpAtlas*; isso corresponde, por exemplo, a uma magnitude limite detectada de $m_g \approx 23$.
- Os objetos com picos múltiplos são dissociados: o pico máximo é marcado como *parent*, e os outros picos como *children*.

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

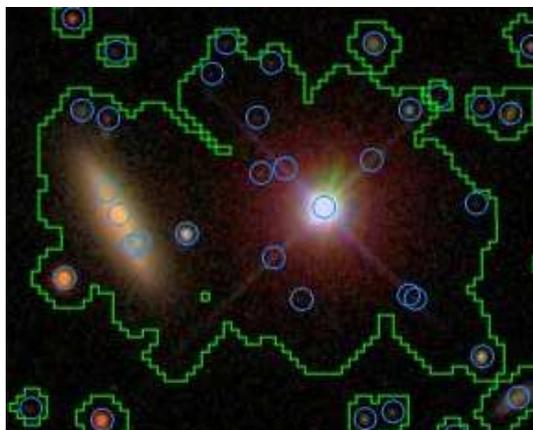


Figura 2.6: A região da galáxia considerada, conforme SDSS. Todos os objetos fotométricos detectados estão circulosados em azul. A linha verde contém todos os objetos children de um mesmo parent.

- Baseado em métodos estatísticos, o *deblender* decide se os objetos fotométricos ao redor de um parent são uma parte dele ou não.

Cada objeto fotométrico detectado durante esse processo é catalogado nos arquivos do SDSS, desconhecendo-se a sua natureza exata. O ponto crítico e mais problemático é a última etapa: como atribuir diferentes fontes de luz a uma única galáxia? Como saber certamente se um pico luminoso, ao estar dentro de um disco galáctico, é um aglomerado estelar, uma estrela do campo ou mesmo um quasar de fundo? Distinguir visualmente a natureza física dos objetos é uma tarefa muito árdua, ainda mais em um processo automatizado.

Os objetos *Fukugita0048* e *Fukugita0049* são etiquetados como *children*, e os algoritmos do SDSS detectam 21 outros irmãos. A Figura 2.6 mostra a região da galáxia considerada, à esquerda, com uma estrela muito brilhante à direita. Todos os objetos fotométricos detectados pelo SDSS estão indicados por um círculo azul. As linhas verdes cercam todos os objetos de uma mesma família, ou seja, que têm o mesmo parent. Assim, o *deblender* do SDSS entende a galáxia, a estrela vizinha e várias outras fontes de luz como um mesmo e único objeto físico. O objeto parent é a estrela brilhante.

Esse exemplo ilustra a complexidade de trabalhar com processos completamente automatizados, bem como a dificuldade de se usar grandes catálogos do SDSS. Antes

de começar a redução dos dados, a escolha dos objetos determinará a viabilidade do estudo.

Quando foi percebido o problema das duas “semigaláxias”, escrevemos para o Prof. Masataka Fukugita, da Universidade de Tóquio, que nos respondeu:

Dear Dr Curty (sic !)

Thank you very much for your mail. Indeed, 48 and 49 are the same galaxy. I know that this happens very often in the original photo catalogue, and we have checked undebled images for each galaxy so that we do not double count galaxies. It seems that what you pointed out is the one escaped from our rejections. Thank you very much for pointing out this.

Since this kind of tests are all eye-based inspections image by image, the result may not be completely error free even if we did checkes as careful as possible. I appreciate your check and informing us.

Yours sincerely,

Masataka Fukugita

Além da abordagem puramente fotométrica, as amostras de galáxias HII são selecionadas por critérios espectroscópicos. A identificação de um espectro como um objeto fotométrico se revela às vezes problemática. Ao confrontar-se com todas essas dificuldades, decidimos trabalhar sobre as imagens do campo do SDSS e, em consequência do que é exposto acima, não usar as segmentações propostas pelo SDSS. Desenvolvemos, então, a nossa própria metodologia para a seleção e a redução dos dados fotométricos.

Nos catálogos de Fukugita (Fukugita et al. 2007) e de Kniazev (Kniazev et al. 2004), as referências dos objetos fotométricos são fornecidas. O catálogo de Izotov (Izotov et al. 2006) contém só as referências dos alvos espectroscópicos, assim tivemos que achar as referências fotométricas com o programa de identificação cruzada (*crossID*), na página web do SDSS. A identificação fotométrica de todos os objetos detectados pelo SDSS é composta por cinco números: *run*, *rerun*, *camcol*, *field* e *obj*. Esses números descrevem vários arquivos, inclusive as imagens de campo (arquivos *fpC*, para *photometric pipeline corrected frames*). As imagens de campo, com o *bias* já subtraído e divididas pelo *flat field*, têm uma dimensão de 2048 x 1489 píxeis,

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

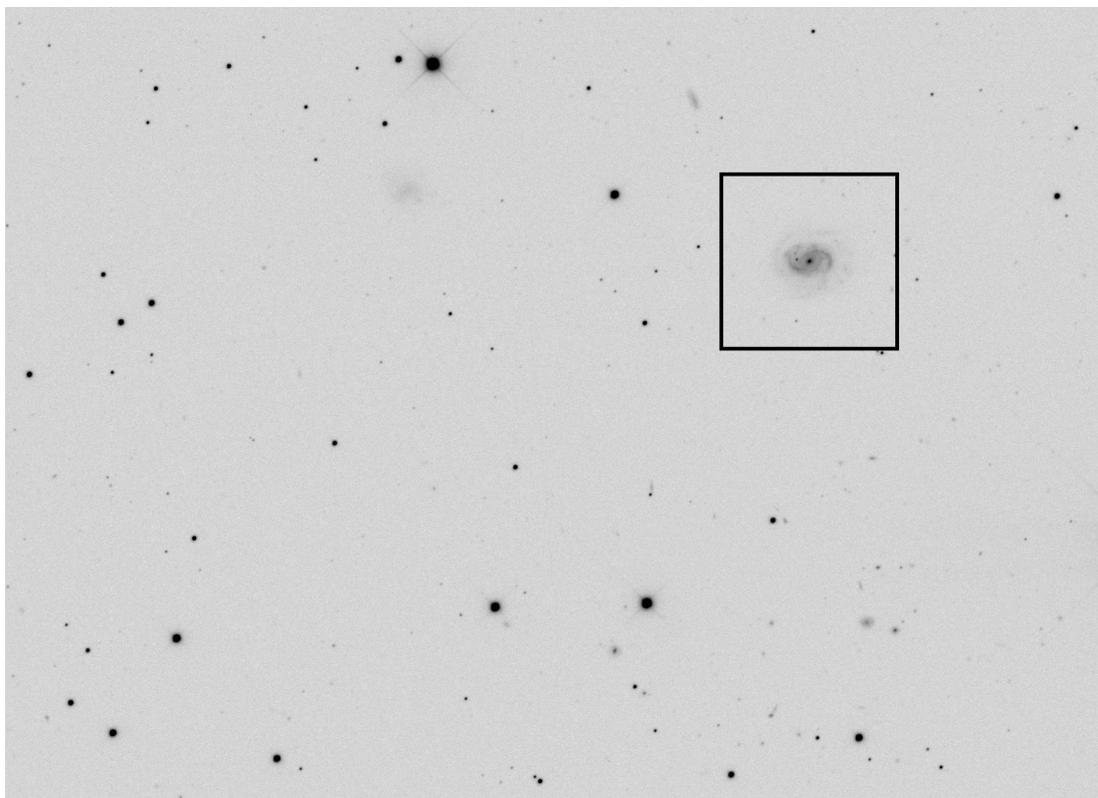


Figura 2.7: Imagem do campo fpC-000752-g4-0284. A galáxia *Fukugita1130* é extraída cortando um quadrado com $6r_{Petro}$ de lado.

com uma escala de $0.396''/\text{pixel}$. Essas imagens não são fotometricamente calibradas, mas as informações necessárias (tempo de exposição (53.907456 s), massa de ar, ponto zero e coeficiente de extinção) ficam nos arquivos `tsField` (*ts* para *target selection*). Um exemplo de imagem de campo do SDSS está na Figura 2.7.

Os arquivos `tsObj` são os catálogos dos objetos fotométricos. Várias grandezas foram calculadas pelo SDSS, dentre elas o raio de Petrosian r_{Petro} (Petrosian 1976). Este foi computado pelo SDSS para todos os objetos fotométricos, também para as estrelas. Usamos esse raio de Petrosian para cortar as imagens do campo a $3r_{Petro}$ ao redor da galáxia.

A imagem assim obtida (Figura 2.8), cortando simplesmente a imagem do campo, não é adequada para fotometria superficial. Os objetos estrangeiros à galáxia, tipicamente as estrelas do campo, podem atrapalhar consideravelmente as medidas. A próxima etapa da redução dos dados consiste, portanto, em selecionar a galáxia desejada

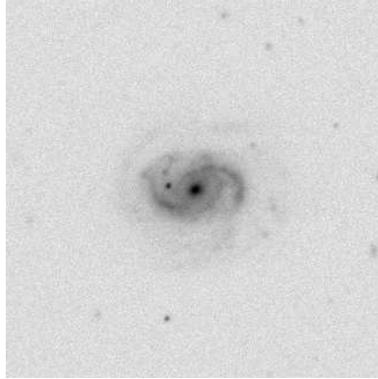


Figura 2.8: A galáxia *Fukugita1130* na banda *g*.

e apagar todos os outros picos de luz.

2.2.2 Limpeza das imagens

Limpamos as imagens para evitar que estrelas ou outros objetos estrangeiros à galáxia atrapalhem as medidas. Os objetos são detectados por um corte em intensidade, ou seja, são identificados como picos luminosos acima de um valor limiar (*threshold*, em inglês). O valor do *threshold* é determinado a partir do histograma da imagem. Como a maioria dos píxeis fazem parte do céu, o histograma caracteriza o ruído da imagem.

O ruído gerado em um detector CCD segue uma distribuição de Poisson. Para uma variável aleatória k , a densidade de probabilidade da lei de Poisson é dada por:

$$p(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad (2.1)$$

onde λ é o parâmetro da lei de Poisson.

A distribuição de Poisson é uma lei de probabilidade discreta, ou seja, definida para valores inteiros e positivos ($k \in \mathbb{N}$). O parâmetro $\lambda > 0$ descreve ao mesmo tempo o valor central, dado pela média \bar{k} , e a dispersão, dada pela variância σ_k^2 . Assim, temos:

$$\bar{k} = \sigma_k^2 = \lambda. \quad (2.2)$$

Para um tratamento matemático mais adequado, é possível estender a lei de Poisson

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

sobre todo o conjunto \mathbb{R} , usando a função gama de Euler Γ :

$$p(x) = \begin{cases} e^{-\lambda} \frac{\lambda^{x-x_0}}{\Gamma(x-x_0+1)} & \text{se } x \geq x_0, \\ 0 & \text{senão.} \end{cases} \quad (2.3)$$

(Para ver um exemplo dessa função, vide a curva vermelha da Figura 2.9.)

O parâmetro complementar x_0 quantifica o deslocamento da função, quando uma constante for adicionada ou subtraída à variável considerada. No caso de uma imagem astronômica, x_0 descreve o nível do céu, de tal maneira que a média se torna nula. À medida que λ aumenta, a função de Poisson toma o aspecto de uma curva de sino e, para grandes valores de λ , tende para a distribuição normal. No entanto, fica ligeiramente assimétrica, o que tem como consequência que o máximo da função – que é, por definição, a moda da distribuição – é um pouco menor que a média. Uma distribuição poissoniana, tal como dada pela equação 2.3, tem as seguintes propriedades:

$$\begin{aligned} \sigma_x^2 &= \lambda \\ \bar{x} &= \lambda - x_0 \end{aligned} \quad (2.4)$$

$$\text{Moda}(x) = \lambda - x_0 - 1/2$$

Para definir o valor do *threshold*, a função de Poisson contínua é ajustada ao histograma, como é feito na Figura 2.9: o histograma da imagem é a linha preta e o ajuste, a vermelha. A moda da distribuição é escolhida como *threshold*.

Um ajuste do histograma é um método robusto para achar uma boa aproximação do nível do céu e do ruído. Para ilustrar isso, o histograma de um pedaço do céu, normalizado para o mesmo número de píxeis, é plotado em cinza na Figura 2.9. O ajuste da função de Poisson fornece um *threshold* ligeiramente menor que o precedente, mostrando que a presença de um objeto, mesmo grande, tem uma influência limitada.

O *threshold* assim determinado é aplicado sobre a imagem original, o que cria uma imagem binária, chamada de máscara: píxeis maiores que o *threshold* assumem o valor 1, os menores, 0. Como o valor do *threshold* é muito próximo do valor do céu, sem suavização⁵ preliminar, a imagem binária é completamente ruidosa. Para aumentar a

⁵A suavização, *smoothing* em inglês, é uma convolução por uma função *kernel* positiva, simétrica e normalizada.

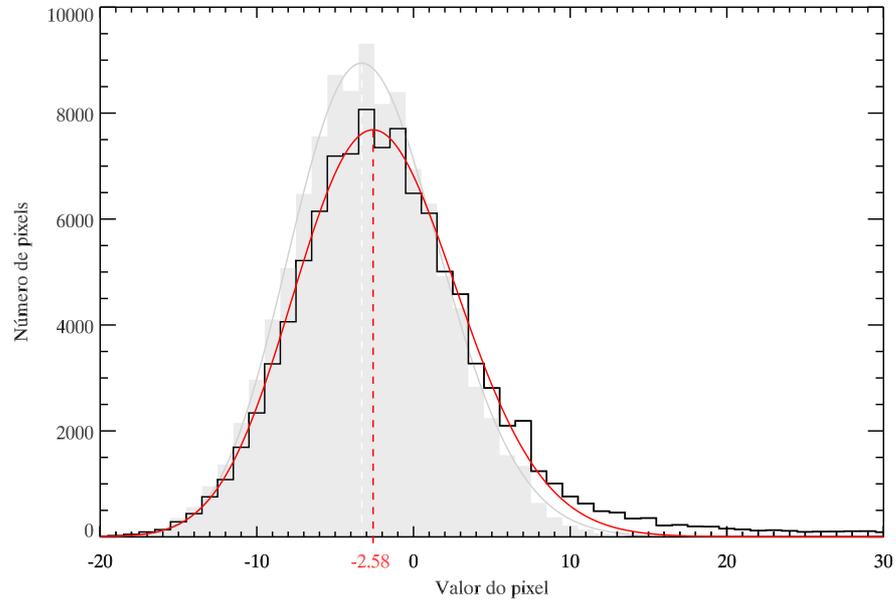


Figura 2.9: Histograma da Imagem 2.8 (linha preta) e o ajuste por uma lei de Poisson (linha vermelha). O máximo da curva determina o *threshold*. O histograma preenchido em cinza é do céu, normalizado com o mesmo número de píxeis.

razão sinal-ruído, dois operadores morfológicos são aplicados consecutivamente, uma erosão seguida de uma dilatação. Os elementos estruturantes são pequenos, de 3×3 píxeis:

$$\text{elemento de erosão} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\text{elemento de dilatação} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Essas operações são semelhantes a uma convolução. A erosão anula os píxeis que não estão cercados de quatro píxeis vizinhos de valor 1, e a dilatação restaura os píxeis perdidos em torno das estruturas sobreviventes. A primeira imagem da Figura 2.10 é

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

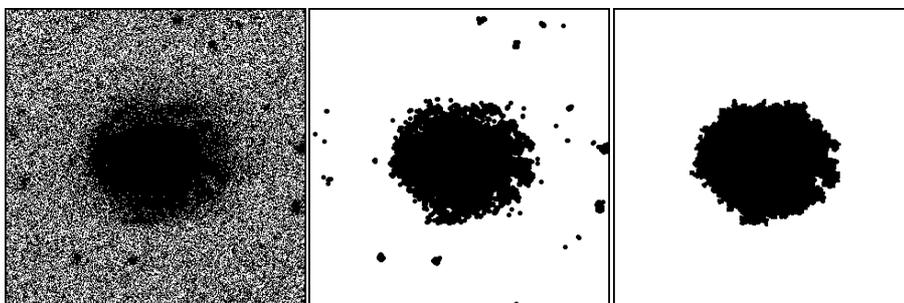


Figura 2.10: Na esquerda, a imagem binária da Figura 2.8: píxeis maiores que o *threshold* estão em branco, menores em preto. No meio, a mesma imagem após a erosão e a dilatação. Na direita, a máscara conserva apenas o objeto central.

a galáxia *Fukugita1130* após o *threshold*. A segunda é a imagem depois da erosão-dilatação. Na última, apenas o objeto central é conservado e os buracos são preenchidos. Esta serve de máscara para a galáxia.

Para afinar a detecção dos objetos ou o limite da galáxia, introduzimos um parâmetro livre para a determinação do *threshold*. O parâmetro livre é definido como a fração do pico da função de Poisson, como indicado na Figura 2.11; o *threshold* é dado pela intensidade onde a função de Poisson cruza o parâmetro livre. Um valor negativo do parâmetro significa que consideramos a parte crescente da função.

Três exemplos de máscaras para a galáxia *Fukugita1130* são representados na Figura 2.12. O mais amplo, obviamente grande demais e incluindo vários objetos externos, foi obtido com o valor do parâmetro livre de -0.5. A máscara do meio foi obtida com o valor padrão de 1, quer dizer que é a mesma máscara que a última da Figura 2.10. A menor máscara resulta de um parâmetro de 0.1 que fornece uma máscara muito pequena, pois não contém a parte externa dos braços espirais.

Tendo as ferramentas adequadas para criar mapas de segmentação com a precisão desejada, podemos definir um procedimento de limpeza das imagens completamente automatizado. A máscara da galáxia é determinada por um primeiro *threshold* da maneira descrita acima. Dentro da máscara, nada é modificado. Fora da máscara, os objetos são detectados com um outro *threshold*. Estas duas etapas são realizadas em simultâneo sobre as cinco bandas *u*, *g*, *r*, *i* e *z*, de tal maneira que um só pixel detectado em uma só banda basta para definir um objeto. As regiões da imagem atribuídas aos objetos detectados são ampliadas proporcionalmente a seus tamanhos. Os píxeis dessas

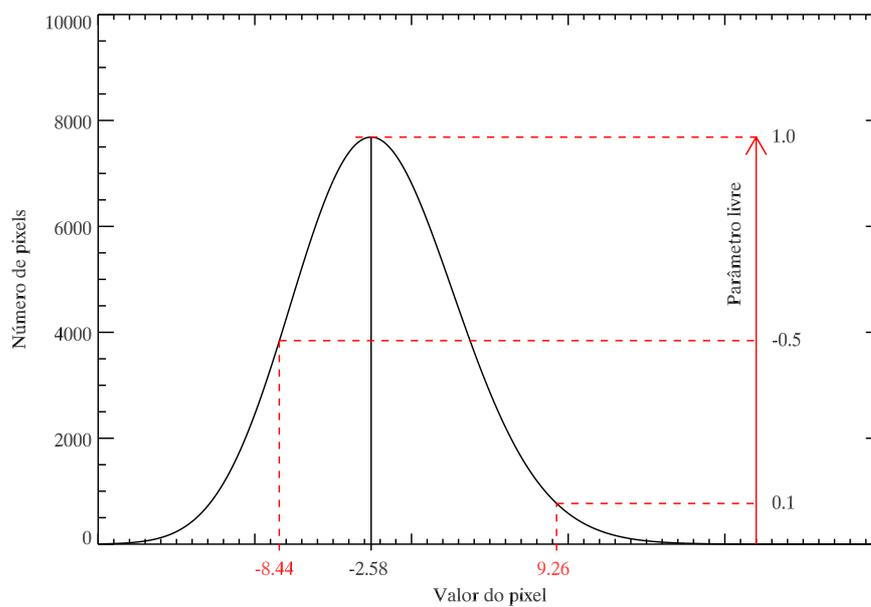


Figura 2.11: Função de Poisson com o parâmetro livre que define o valor do *threshold*.

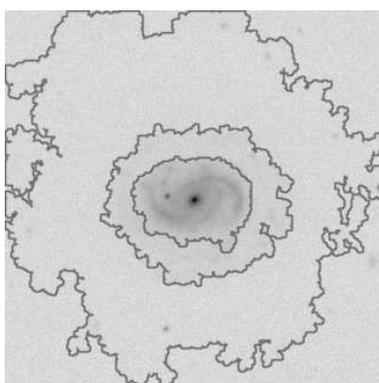


Figura 2.12: A galáxia *Fukugita1130* com os contornos das três máscaras resultantes dos três *thresholds* da Figura 2.11; o mais baixo (-8.44) é o mais externo.

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

regiões são substituídos por píxeis aleatórios cuja distribuição de Poisson foi ajustada no histograma do céu.

O processo é repetido três vezes com valores diferentes para o parâmetro livre:

Primeira limpeza: `galaxylevel=-0.7, starslevel=.5`

Segunda limpeza: `galaxylevel=-0.8, starslevel=.4`

Terceira limpeza: `galaxylevel= 1.0, starslevel=.01`

À medida que a máscara da galáxia se torna menor, o critério de detecção dos objetos fica mais restritivo. Isso para evitar que a parte externa da galáxia, com um baixo brilho superficial, seja sistematicamente substituída por outros píxeis.

A Figura 2.13 mostra o procedimento de limpeza para a galáxia *Fukugita1130*, etapa por etapa. Na coluna da direita, a área da imagem atribuída à galáxia está desenhada mais clara e os objetos detectados estão circulos. Já na segunda limpeza, poucos objetos são detectados, e todos são pequenos. Na terceira limpeza, não é detectado mais nenhum objeto. Na última linha, a imagem limpa é aquela sobre a qual todas as grandezas fotométricas e morfológicas serão computadas. À direita, estão desenhados todos os objetos apagados durante o procedimento.

Sendo o procedimento de limpeza inteiramente automatizado, é possível que apareçam alguns erros em cada uma das diferentes etapas. Em particular, podem aparecer absurdos quando a galáxia não se encontra no centro da imagem ou se há, na imagem, um objeto maior do que a própria galáxia. Além disso, o problema de sobreposição não é resolvido quando o objeto estrangeiro se encontra dentro da máscara da galáxia. Mas, mesmo assim, na grande maioria dos casos, o resultado da limpeza é satisfatório. Por comparação, o mapa de segmentação fornecido pelo SDSS é semelhante ao nosso (Figura 2.14).

Decidimos não mexer em nada dentro da máscara da galáxia. Isso pode ter consequências drásticas quando há realmente uma estrela do primeiro plano no meio da galáxia, mas evita também que sejam apagadas partes ou elementos pertencentes à galáxia, como braços espirais, aglomerados gigantes ou regiões HII. Além disso, é muito difícil discernir uma estrela do campo com baixa luminosidade de um elemento da galáxia, uma região HII, por exemplo. Portanto, preferimos deixar as galáxias tais como estão ao final do procedimento de limpeza, conscientes de que a presença de uma estrela em algumas imagens pode falsear as medidas fotométricas.

2.2 Redução dos dados fotométricos

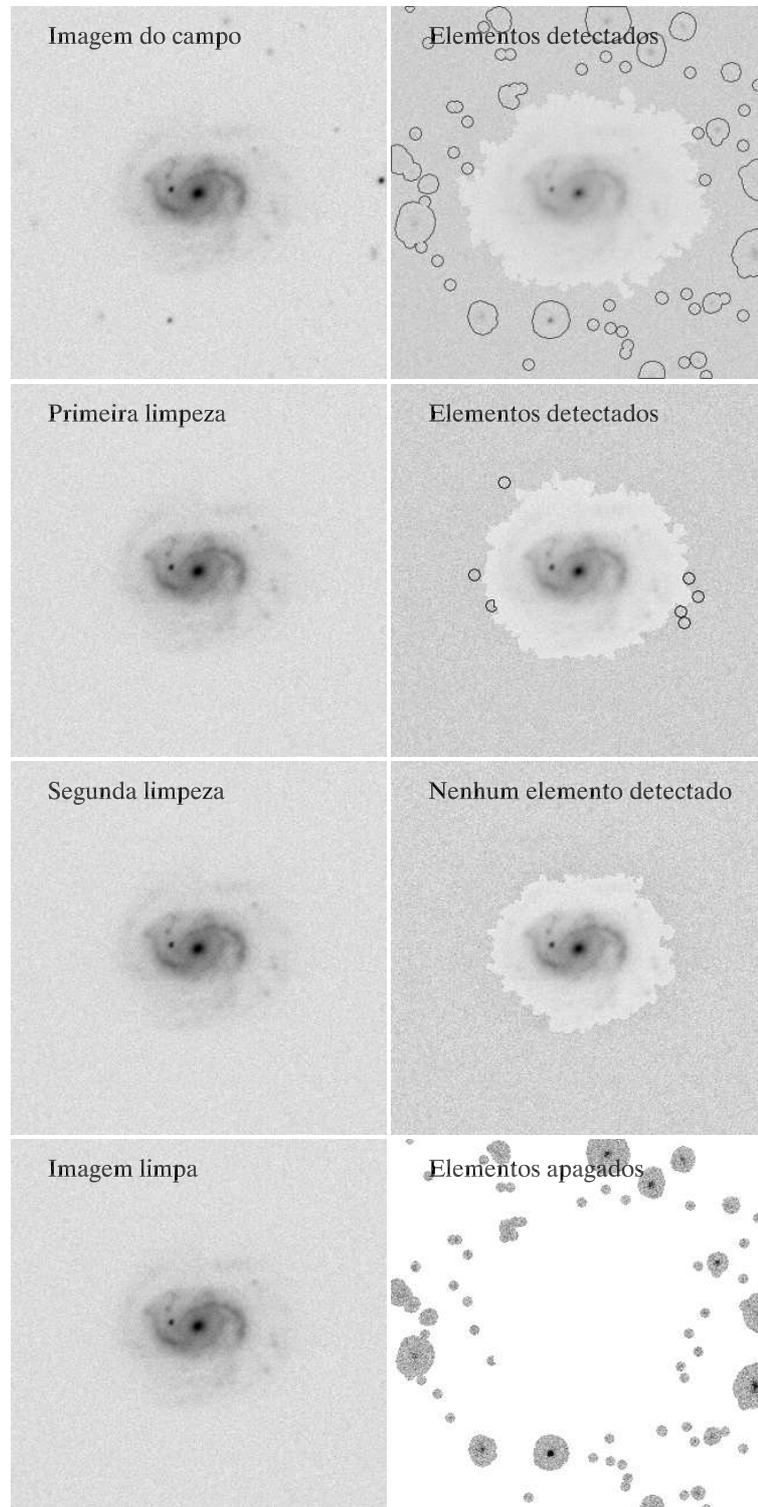


Figura 2.13: Procedimento completo de limpeza das imagens, exemplo com *Fukugita1130*. Na última linha, a imagem limpa e todos os objetos removidos.

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

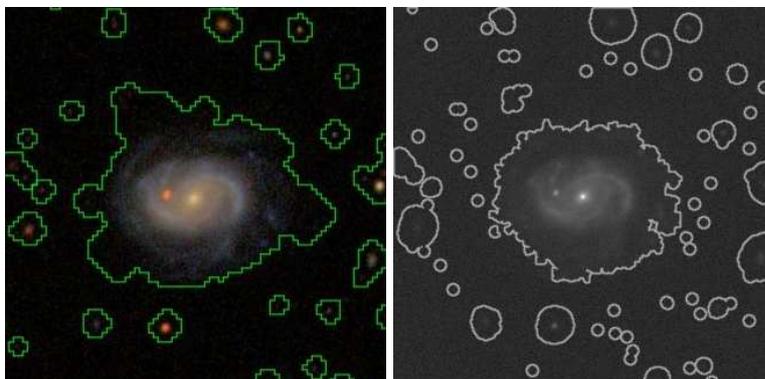


Figura 2.14: Mapa de segmentação computado pelo SDSS (esquerda) e aquele que decorre do nosso procedimento de limpeza (direita).

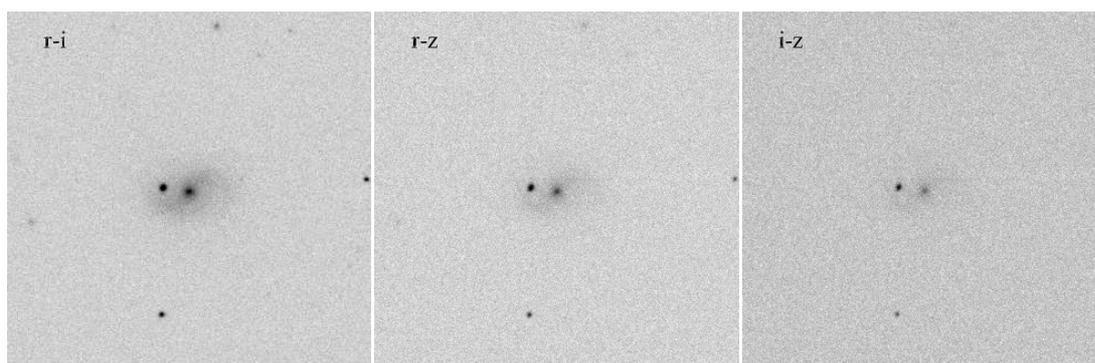


Figura 2.15: Imagem em $r-i$, $r-z$, $i-z$ da galáxia *Fukugita1130*.

Contudo, podemos sugerir umas ideias como pistas para melhorar a detecção de estrelas sobrepostas a uma galáxia. O método usual para identificar estrelas é usar o perfil de luminosidade. Mas dificilmente é aplicável para diferenciar estrelas de regiões HII, pois pequenos objetos têm, todos, o mesmo perfil, o da PSF (*point spread function*).

Poderíamos, então, usar cores, trabalhando sobre imagens obtidas pela diferença de dois filtros. As estrelas do campo são frequentemente velhas e vermelhas, enquanto as regiões HII são azuis. Uma imagem construída pela subtração de bandas vermelhas e infravermelhas amplifica as estrelas e atenua os braços espirais, como pode ser visto na Figura 2.15.

No entanto, a cor não é uma panaceia! Uma estrela *early-type*, mais quente, seria

difícil de detectar “em cima” de uma galáxia *late-type*, e vice-versa. Uma estrela fria poderia ser confundida com um aglomerado estelar.

O nosso método de limpeza de imagem é adaptado para a grande maioria das galáxias estudadas em nossos catálogos. Mas certas galáxias ainda ficam “poluídas” por estrelas do campo ou galáxias vizinhas. Alguns exemplos podem ser vistos nas Figuras 6 e 11 do Capítulo 5.

2.2.3 Subtração do céu

Uma última operação é indispensável antes de iniciar as medidas: o nível do céu deve ser subtraído da imagem. A subtração do céu é uma fase muito importante, porque se não for realizada com precisão, ele ainda contribui com o fluxo e todas as medidas de intensidades são sobrevalorizadas.

Para cumprir a subtração do céu, uma função do primeiro grau, ou seja, um plano, é ajustada às intensidades dos píxeis do céu. O ponto delicado é a seleção da área apropriada para evitar incluir no céu uma parte do fluxo externo da galáxia. Efetivamente, as galáxias elípticas, seguindo a lei de de Vaucouleurs, são espacialmente muito estendidas e uma fração significativa do fluxo poderia ser atribuída ao céu. Isso teria como consequência a subtração de um nível de céu elevado demais e, possivelmente, medir fluxos integrados negativos⁶.

A fim de evitar isso, é intencionalmente escolhida uma máscara de galáxia muito grande. Primeiro, uma máscara é determinada por um *threshold* com um parâmetro livre `galaxylevel=-0.7`, o que corresponde mais ou menos à primeira máscara do procedimento de limpeza, no canto superior direito da Figura 2.13. Em seguida, ela é dilatada por um elemento circular de raio proporcional ao tamanho da máscara.

O processo de subtração do céu é resumido na Figura 2.16. Devido à necessidade de uma representação sóbria e entendível, o problema bidimensional foi simplificado em uma dimensão: é desenhado somente um corte da imagem ao longo do eixo X e passando pelo centro galáctico. A galáxia se encontra em cinza, e o céu em preto, com as retas verticais mostrando os limites das máscaras. A reta vermelha é o resultado

⁶Corolário: pode-se encontrar certas situações muito desagradáveis, como curvas de crescimento decrescentes ou raios galácticos negativos ou mesmo imaginários!

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

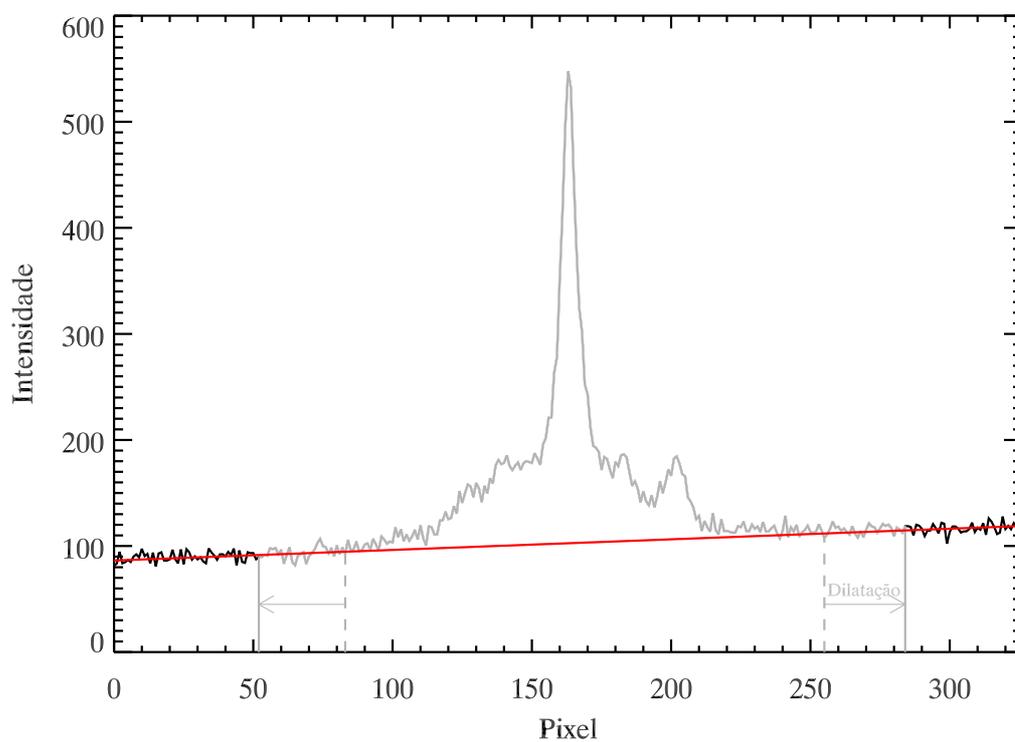


Figura 2.16: Corte ao longo do eixo X da galáxia *Fukugita1130*. A linha cinza corresponde aos píxeis da galáxia, enquanto a preta corresponde ao céu. As retas verticais são os limites da máscara da galáxia após (linha contínua) e antes (linha tracejada) a dilatação. A linha vermelha é o nível do céu, resultante de um ajuste linear da curva preta. A inclinação do céu é exagerada para maior clareza.

de uma regressão linear múltipla que inclui apenas os píxeis do céu. É essa função vermelha que é subtraída da imagem.

Concluindo, o nosso procedimento completo de redução dos dados envolve as seguintes etapas: a extração das galáxias a partir das imagens do campo, a limpeza das imagens e a subtração do céu. Essas três operações foram aplicadas sobre todas as galáxias entrando nos estudos morfológicos, ou seja, aquelas dos catálogos de Fukugita, Kniazev e Izotov. Todos os processos são inteiramente automatizados. Os parâmetros livres de entrada dos diferentes procedimentos computacionais são sempre determinados de maneira autônoma, consoante as propriedades da imagem.

Sem controle visual sistemático, custoso para mais de 3000 galáxias, temos que aceitar o risco de encontrar alguns dados mal reduzidos. Por conseguinte, devemos prestar atenção, em nossas análises estatísticas, à presença de medidas inconsistentes criando possíveis pontos divergentes.

2.3 Abertura

As galáxias não são fontes pontuais. Precisa-se definir quais píxeis da imagem são atribuídos à galáxia, e quais ao céu. Essa etapa é crucial, porque todas as medidas fotométricas serão computadas dentro da abertura. Uma abertura mal definida pode introduzir um viés. Se a abertura for pequena demais, os cálculos serão realizados somente na área interna da galáxia, mas se a abertura for grande demais, uma porção importante do céu será considerada como parte da galáxia. Geralmente, executa-se os seguintes passos para criar uma máscara adequada:

1. Detecção da galáxia e primeira estimativa da máscara pelo método do *threshold*.
2. Nessa primeira máscara, cálculo do centro de luz (momento de primeira ordem).
3. Nessa primeira máscara, cálculo da elipse que melhor descreva a galáxia (a partir do momento de segunda ordem).
4. Com aberturas elípticas, cálculo de um raio característico, tal como o raio de Kron ou o raio de Petrosian⁷.
5. Determinação da máscara final baseada no raio característico.

⁷É claro que é possível calcular o raio característico usando aberturas circulares mas, obviamente, se perde em precisão, sobretudo para as galáxias *edge-on*.

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

O raio característico deve ser definido de maneira intrínseca à galáxia, sem referência ao ruído na imagem. É a única maneira aceitável de definir a fronteira da galáxia, porque ela é pouco influenciada pelas condições de observação, idealmente independente delas.

2.3.1 Aproximação com um *threshold*

A abertura mais fácil e mais intuitiva é aquela definida por uma máscara baseada em um *threshold*. Tipicamente, se escolhe um *threshold* a 3σ , isso quer dizer que sobrevivem ao corte só os píxeis com uma intensidade maior do que três vezes o ruído – definido como o desvio padrão dos píxeis do céu – acima do nível do céu. O conjunto dos píxeis contíguos maiores do que o *threshold* define a máscara da galáxia. Esse processo cria uma isofota, cuja magnitude não é fixada de maneira absoluta, mas relativamente ao ruído da imagem.

Um *threshold* a 3σ é com frequência usado para detectar os objetos numa imagem. Mas para conter a totalidade da galáxia, o valor de *threshold* deve ser muito mais perto do nível do céu. No entanto, é justamente neste caso que o ruído conduz a isofotas irregulares. Pode-se reduzir o ruído por um suavização da imagem. Mas esta tem como inconveniente espalhar o fluxo sobre uma área maior e arredondada e, por consequência, criar máscaras maiores e mais circulares. Para evitar esse efeito indesejável, preferimos usar operadores morfológicos, como a dilatação e a erosão apresentadas na secção 2.2.2.

Uma máscara construída com um *threshold* pode ser usada para trabalhos intermediários, como fizemos na redução dos dados. Porém, esse método não é rigoroso o suficiente para definir com precisão a área da galáxia, porque depende muito da qualidade da imagem, particularmente a razão sinal/ruído ou a subtração do céu. De início, tínhamos realizado as nossas medições fotométricas dentro da máscara feita por um *threshold*. Mas observamos uma propriedade estranha: a razão sinal/ruído aumentava junto com o *redshift*! Obviamente, isso não é um comportamento real, mas, ao contrário, um viés induzido pela escolha de uma abertura inadequada.

2.3.2 Os momentos de uma imagem

Quando uma primeira máscara aproxima-se da posição e do tamanho da galáxia, pode-se calcular o centro de luz. Semelhante ao baricentro para massa, o centro de luz é definido como a posição média da galáxia em ambas as dimensões X e Y da imagem. Isso é, por definição, o momento de primeira ordem.

Em matemática, um momento M_k de ordem k de uma função $f(r)$ é definido como:

$$M_k(f) = \frac{1}{f_{tot}} \int_E r^k f(r) dr. \quad (2.5)$$

A variável r é um elemento do espaço vetorial E . O primeiro fator da equação 2.5 é a normalização da função f :

$$f_{tot} = \int_E f(r) dr. \quad (2.6)$$

Em cálculo de probabilidade, f é a densidade de probabilidade da variável aleatória r . Nesse caso, a integral de $f(r)$ sobre todo o espaço E é igual à unidade.

É importante notar que, no caso geral de um vetor aleatório r de dimensão n , os momentos M_k são tensores de ordem k de mesma dimensão do que a do vetor. Consequentemente, temos:

- M_0 é um escalar; sem normalização, M_0 é o total da função f ; com a normalização na definição 2.5, M_0 é sempre igual a 1.

- M_1 é um vetor; é a média das variáveis r .

$$M_1 = \langle r \rangle = \begin{pmatrix} \langle r_1 \rangle \\ \langle r_2 \rangle \\ \dots \\ \langle r_n \rangle \end{pmatrix}.$$

- M_2 é uma matriz simétrica definida positiva; se as variáveis são centradas, ou seja, se as suas médias foram subtraídas, M_2 é a matriz da covariância⁸ das variáveis r .

⁸A matriz da covariância é, às vezes, chamada de matriz da variância-covariância, porque a sua diagonal é formada das variâncias. A variância é o quadrado do desvio padrão σ .

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

$$M_2 = \text{Cov}(r) = \begin{pmatrix} \sigma_{r_1}^2 & \sigma_{r_1}\sigma_{r_2} & \dots & \sigma_{r_1}\sigma_{r_n} \\ \sigma_{r_1}\sigma_{r_2} & \sigma_{r_2}^2 & \dots & \sigma_{r_2}\sigma_{r_n} \\ \dots & \dots & \dots & \dots \\ \sigma_{r_1}\sigma_{r_n} & \sigma_{r_2}\sigma_{r_n} & \dots & \sigma_{r_n}^2 \end{pmatrix}.$$

- M_3 é um tensor de terceira ordem; é relacionado com as obliquidades ou assimetrias (*skewness*, em inglês) das variáveis r .
- M_4 é um tensor de quarta ordem; é relacionado com as curtoses (*kurtosis*, em inglês) das variáveis r .

Uma imagem de m linhas por n colunas pode ser descrita como uma função discreta $I(x, y)$ de duas variáveis $x = 1, 2, \dots, n$ e $y = 1, 2, \dots, m$ indicando a posição do pixel:

$$I(x, y) = \begin{pmatrix} I_{11} & I_{12} & \dots & I_{1n} \\ I_{21} & I_{22} & \dots & I_{2n} \\ \dots & \dots & \dots & \dots \\ I_{m1} & I_{m2} & \dots & I_{mn} \end{pmatrix}. \quad (2.7)$$

Os momentos de uma imagem são portanto calculados da seguinte maneira:

$$M_k(I) = \frac{1}{I_{tot}} \sum_{x=1}^m \sum_{y=1}^n \binom{x}{y}^k I(x, y). \quad (2.8)$$

Os momentos de segunda ordem definem as dispersões estatísticas da variável aleatória em torno da sua posição média (através das variâncias), assim como das variações simultâneas de cada par de variáveis (através das covariâncias). No caso de uma imagem, os segundos momentos são uma medida da distribuição de luz ao longo das direções X e Y , mas também na direção transversal XY da diagonal. Essas informações podem ser usadas para descrever uma galáxia como uma elipse.

Conhecendo o vetor M_1 e a matriz M_2 , é possível modelar a distribuição da luz na imagem I por uma gaussiana bidimensional (veja, por exemplo, *Source Extractor for Dummies* Holwerda 2005). A abordagem é exatamente a mesma que a determinação dos componentes principais de uma nuvem de pontos, como descrito no Apêndice A. Quando a gaussiana é uma boa aproximação da distribuição espacial da intensidade da galáxia, a quase totalidade da luz irá cair dentro da elipse que se estende até 3σ

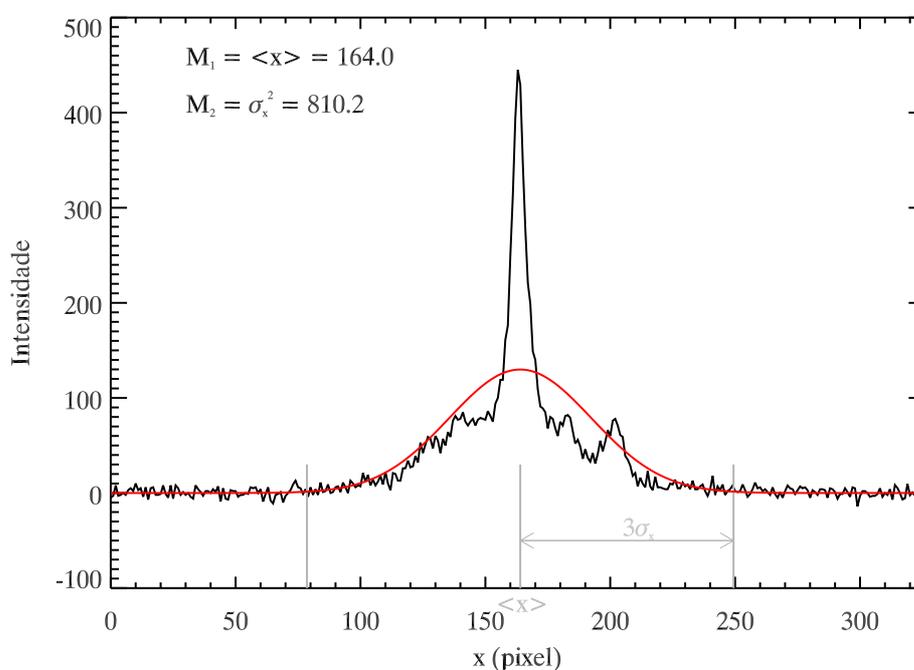


Figura 2.17: Corte da galáxia *Fukugita1130* ao longo do eixo X (linha preta). Calcula-se os momentos de 1^a e 2^a ordens para definir a gaussiana (linha vermelha). A borda a 3σ da gaussiana define uma elipse que pode ser usada como abertura da galáxia.

em cada um dos dois eixos da gaussiana⁹. Ilustramos um exemplo unidimensional na Figura 2.17.

Os momentos de primeira ordem são o centro de luz da galáxia e definem o centro da gaussiana. Os momentos de segunda ordem definem os eixos da gaussiana, ou seja, a posição angular e a alongação¹⁰. Essas propriedades geométricas permitem melhorar a fotometria de abertura, usando aberturas elípticas em vez de circulares, para traçar a curva de crescimento da galáxia, por exemplo. Neste trabalho, usamos sempre aberturas elípticas para a fotometria de abertura.

A elipse delimitando a gaussiana até uma dispersão de 3σ pode ser empregada

⁹Atenção: O σ usado até agora estava definido como o desvio padrão da intensidade do céu. O σ aqui é o parâmetro da definição da gaussiana; este σ não é uma intensidade mas um comprimento, expresso em píxeis.

¹⁰Se A e B são os eixos de uma elipse, A/B é chamado de alongação e $1 - B/A$ é chamado de elipticidade. Mais conhecido, $\sqrt{1 - B^2/A^2}$ é a excentricidade.

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

como máscara da galáxia. Mas, na literatura, os momentos de segunda ordem nunca são usados como dimensões características da galáxia. Isso se explica pelo fato de que todo o desenvolvimento analítico ao redor dos momentos é baseado na hipótese de que uma galáxia é bem descrita por uma função gaussiana. Mas nenhuma galáxia apresenta um perfil gaussiano: as galáxias elípticas seguem a lei de de Vaucouleurs e os discos galácticos mostram um perfil exponencial. Precisa-se, então, de uma outra definição de raio característico, sem referência a uma lei subjacente.

2.3.3 O raio de Kron

Uma possibilidade de dimensão característica é usar o raio de Kron (Kron 1980), definido como o momento de primeira ordem do raio:

$$\begin{aligned} r_{Kron} &= \frac{\sum_i r_i I_i}{\sum_i I_i} \\ &= \frac{\sum_x \sum_y \sqrt{(x - \bar{x})^2 + (y - \bar{y})^2} I(x, y)}{\sum_x \sum_y I(x, y)}. \end{aligned} \quad (2.9)$$

O momento de primeira ordem calcula a média de uma variável. Então, o raio de Kron pode ser visto como o raio médio da galáxia. Kron (1980) indica que a maioria do fluxo (> 90%) é contido em uma abertura de $2r_{Kron}$ e que seu raio característico é válido independentemente do tipo galáctico. No começo definido com uma abertura circular, o raio de Kron pode ser adaptado para aberturas elípticas. A máscara padrão de SExtractor (Source-Extractor, Bertin & Arnouts 1996), por exemplo, é uma elipse, cujos eixos são $2.5r_{Kron}$ corrigidos da elongação.

2.3.4 O raio de Petrosian

Uma outra possibilidade de dimensão característica é usar o raio de Petrosian (Petrosian 1976). O raio de Petrosian é a distância r_{Petro} do centro onde o brilho superficial μ vale uma fração – geralmente 0.2 – do brilho superficial médio no interior de r_{Petro} :

$$\frac{\langle \mu(r = r_{Petro}) \rangle}{\langle \mu(r < r_{Petro}) \rangle} = 0.2. \quad (2.10)$$

Concretamente, para calcular o raio de Petrosian, avalia-se a curva de luz, ou seja, a distribuição radial da luz. A curva de luz é obtida medindo o fluxo em anéis sucessivos com o mesmo centro que a luz da galáxia. O raio de Petrosian com uma abertura circular é calculado no catálogo do SDSS para todos os objetos fotométricos.

2.3.5 A isofota de Petrosian

Uma máscara definida com uma isofota tem a vantagem de conservar a forma potencialmente irregular da galáxia, mas tem a desvantagem da escolha arbitrária de uma magnitude. Uma máscara calculada com uma dimensão característica da galáxia, como o raio de Kron ou de Petrosian, tem a conveniência de usar uma medida intrínseca à galáxia, mas em contrapartida assume *a priori* uma forma circular ou elíptica para a galáxia.

A ideia é combinar as duas vantagens em uma nova máscara sendo a “isofota de Petrosian”. O valor do *threshold* é escolhido de tal maneira que o brilho superficial sobre a isofota vale um quinto do brilho superficial médio em seu interior. No lugar de um raio específico, obtemos uma intensidade específica I_{Petro} . A equação 2.10 se torna:

$$\frac{\mu(I = I_{Petro})}{\langle \mu(I > I_{Petro}) \rangle} = 0.2 . \quad (2.11)$$

Para estabelecer o valor do *threshold* I_{Petro} , calcula-se a razão da equação 2.11 para uma série de intensidades até obter o valor esperado de 0.2. Como a isofota de Petrosian tem uma intensidade próxima daquela do céu, é frequentemente perturbada pelo ruído presente na imagem. Assim, é necessário filtrar a imagem antes de efetuar os *thresholds*.

Escolhemos uma filtragem em frequência. Aplicamos, como filtro passa-baixo, a função de Butterworth (Figura 2.18):

$$B(\nu) = \frac{1}{\sqrt{1 + (\nu/\nu_c)^2}}, \quad (2.12)$$

onde ν_c é a frequência de corte (*cutoff*, em inglês).

O filtro passa-baixo é aplicado na imagem no espaço das frequências. A transformada de Fourier da imagem é, portanto, multiplicada pelo filtro $B(\nu)$, antes que a

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

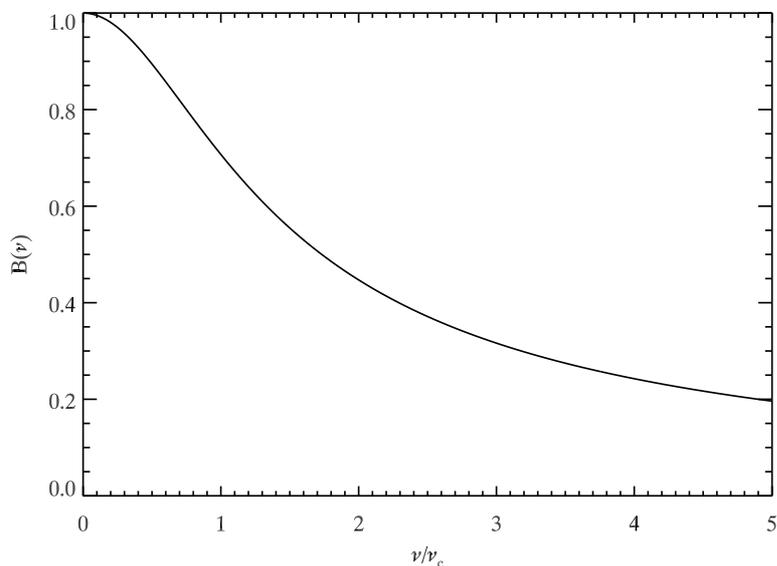


Figura 2.18: Função de Butterworth usada como filtro passa-baixo na imagem antes de computar a isofota de Petrosian.

imagem seja restaurada pela transformada de Fourier inversa. Registra-se que chamar ν de frequência é um abuso de linguagem no caso de tratamento de imagem; uma imagem não é um sinal temporal. É por isso que ν é às vezes chamado de frequência espacial e o quadrado da transformada de Fourier, de espectro de potência espacial.

A imagem doravante filtrada contém menos ruído de alta frequência e a isofota de Petrosian pode ser computada. No entanto, a isofota permanece bagunçada pelo ruído de baixa frequência. Por isso, uma segunda filtragem é aplicada sobre a imagem binária da isofota de Petrosian. Dessa vez, escolhamos uma suavização com uma gaussiana¹¹ cujo desvio padrão é proporcional ($0.3\times$) aos eixos da elipse a 3σ , conforme os momentos de segunda ordem.

Resumindo, apresentamos cinco máscaras possíveis para a galáxia: um *threshold*, a elipse dos segundos momentos, o raio de Kron, o raio de Petrosian e a isofota de Petrosian. Na Figura 2.19, todas as cinco aberturas desenvolvidas nesta seção são re-

¹¹A suavização é uma convolução. Ora, um produto de convolução é equivalente ao produto usual no espaço de Fourier. Realizar uma suavização por uma gaussiana equivale a multiplicar a transformada de Fourier do sinal com a transformada de Fourier da gaussiana, sendo ela mesma uma gaussiana de desvio padrão $1/\sigma$.

presentadas. A primeira máscara, no recanto superior direito, é um *threshold* com o valor do máximo do histograma da imagem. Esta máscara não pode ser usada na fotometria superficial, porque depende do ruído. Todavia, sempre intervém na redução dos dados, e também como primeira aproximação para calcular os momentos da imagem. A segunda máscara é a elipse calculada dos momentos da imagem. A geometria da elipse assim definida é usada cada vez que se trata de fotometria de abertura: raio de Kron, raio de Petrosian, curva de crescimento. Os “raios” de Kron e de Petrosian, na segunda linha da Figura 2.19, são dimensões características da galáxia, cuja medida é afinada por aberturas elípticas. Finalmente, a última máscara, no recanto inferior direito, é a isofota de Petrosian. De concepção original, desenvolvida com a nossa própria metodologia, a isofota de Petrosian é a única máscara que conjuga as vantagens de uma isofota e da qualidade intrínseca do raio de Petrosian.

Em todo nosso trabalho, a isofota de Petrosian define onde começa e termina a galáxia: os píxeis de dentro são atribuídos à galáxia e aqueles de fora, ao céu. Sempre que não haja indicação contrária, esta máscara é usada nas computações das grandezas morfológicas e fotométricas. Por exemplo, considera-se como fluxo total de uma galáxia o fluxo dentro da isofota de Petrosian. Frequentemente, a imagem é multiplicada pela máscara antes de iniciar qualquer cálculo.

2. DADOS E PROCESSAMENTO PRELIMINAR DE IMAGEM

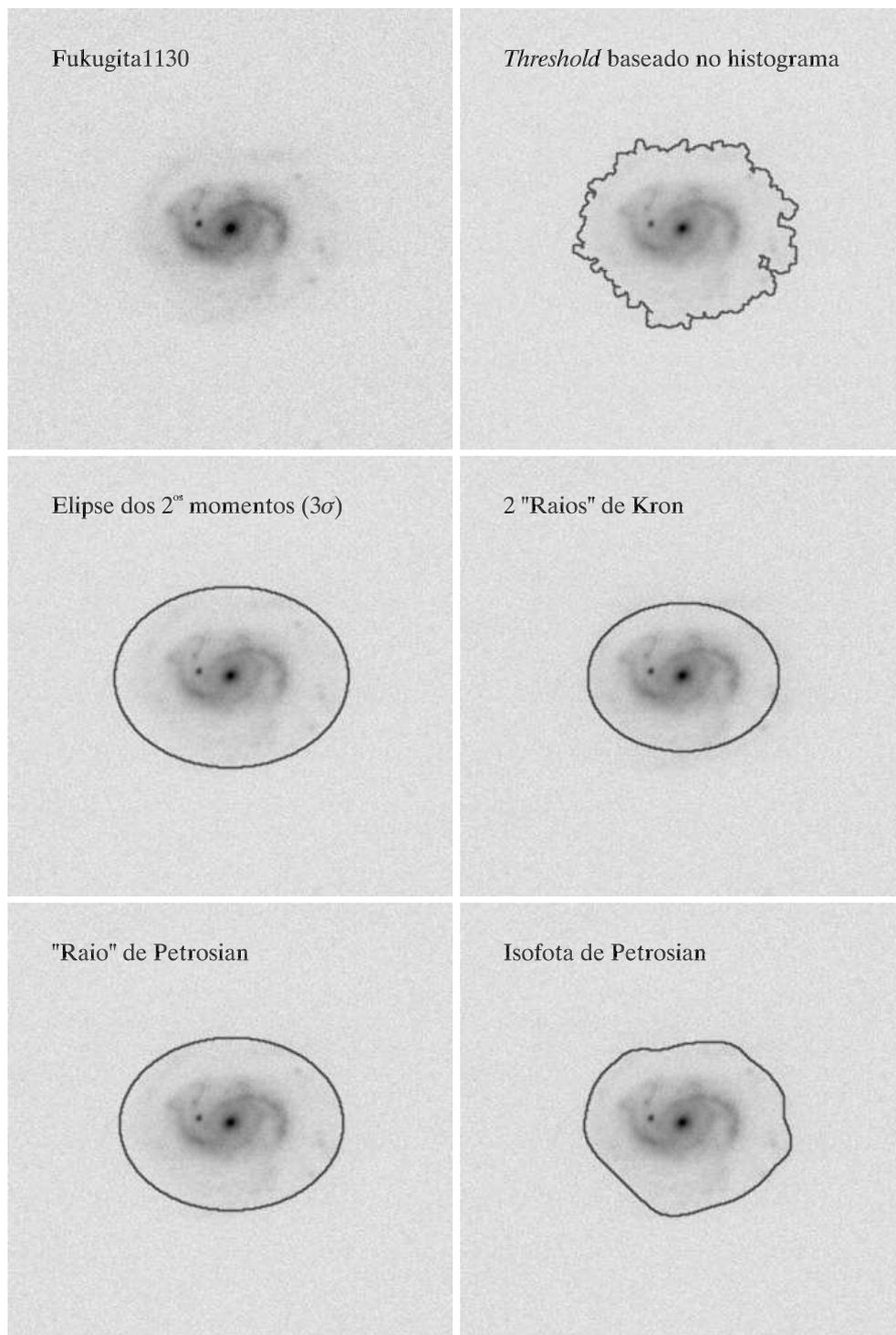


Figura 2.19: Todas as aberturas desenvolvidas em nosso trabalho. Usamos a isofota de Petrosian para as medidas fotométricas.

Capítulo 3

Morfologia quantitativa

Em princípio, e conforme a sua etimologia, a morfologia é uma qualidade, e não uma quantidade. Mas o nosso propósito é justamente medir a morfologia ou, pelo menos, caracterizá-la por grandezas mensuráveis. Somos, portanto, levados a definir de modo arbitrária uma certa quantificação ou modelização da morfologia, o que passa pela adoção de tipos morfológicos. A morfologia da maioria das galáxias no universo local se encaixa na sequência de Hubble (Hubble 1926). Para criar uma classificação simples, parece natural ordenar unidimensionalmente os tipos galácticos da seguinte maneira: elípticas – lenticulares – espirais – irregulares.

A determinação do tipo morfológico, embora simplificado para um só número, é uma tarefa complexa, porque envolve vários aspectos, tais como a forma geral, a presença de estruturas mais ou menos visíveis e organizadas (braços espirais, bojo, barra, anel), além da textura que a galáxia parece ter, às vezes chamada de flokulência. A morfologia tem que integrar todas estas propriedades num único conceito. Essa necessária síntese é tão abrangente que, mesmo na era dos grandes levantamentos de objetos celestes, certas classificações recorrem a um instrumento secular e menos arcaico do que parece: o olho humano.

A carga enorme de trabalho para classificar visualmente o número crescente de galáxias observadas fez nascer algo chamado “ciência cidadã”. Propondo a todos aqueles que queiram classificar galáxias na internet, o projeto *Galaxy Zoo* (Fortson et al. 2011; Lintott et al. 2008) utiliza em grande escala a capacidade de reconhecimento, de forma que possui a visão humana. Os primeiros resultados contêm um catálogo de quase

3. MORFOLOGIA QUANTITATIVA

900 000 galáxias, cujas morfologias foram estimadas por centenas de milhares de voluntários (Lintott et al. 2011).

A alternativa às classificações visuais é recorrer à análise computadorizada. Ora, para tentar reproduzir a sequência de Hubble – ou qualquer outra classificação – de forma automatizada, é necessário definir critérios mensuráveis, chamados índices morfológicos. Este é o assunto da fotometria superficial, que consiste em analisar a distribuição do brilho superficial por diferentes técnicas de processamento de imagem.

Na literatura, vários índices morfológicos foram desenvolvidos durante as últimas décadas. Aos índices comuns (Seção 3.2), juntamos dois novos (Seção 3.3) que definimos especialmente para estudar galáxias em interação: o índice de anelipticidade α e o índice de granulação γ .

As definições formais e as expressões matemáticas dos índices morfológicos são resumidas no Capítulo 5. Assim, não será raro, para questões específicas, que o referamos ao leitor. No capítulo presente, adotamos uma abordagem mais intuitiva, tentando explicar as ideias por trás dos cálculos. Através de três exemplos de galáxias de tipos diferentes, *Fukugita1849* de tipo E ($T=0$), *Fukugita0443* de tipo Sc ($T=4$) e *Fukugita0022* de tipo Im ($T=6$), ilustramos as etapas do processo de computação.

Todas as grandezas morfológicas são computadas sobre as imagens em banda g , que é a mais próxima da banda V do tradicional sistema de Johnson (UBV). Essa escolha foi guiada pelo fato de que a classificação visual de Fukugita et al. (2007) foi realizada na banda g . A vontade de trabalhar com uma só banda vem de uma certa ortodoxia: a morfologia é definida unicamente pela aparência, as cores ou as diferenças entre os filtros já são propriedades físicas.

Os índices morfológicos são calculados para todas as galáxias do catálogo de Fukugita. Como já conhecemos os tipos morfológicos, é possível, e com frequência mais interessante, analisar o comportamento global dos índices de acordo com os tipos, e não de acordo com cada galáxia individualmente. Portanto, como apresentado no Apêndice A, consideramos, às vezes, apenas o valor central (geralmente, a mediana) e uma indicação da dispersão (o desvio absoluto mediano) dos nossos resultados.

A robustez estatística dos índices e a influência das condições observacionais são estudadas no Capítulo 5, Seção 4.3. Analisamos aqui a capacidade de cada índice de separar galáxias *early-type* de *late-type*. Consideramos como galáxias *early-type* as elípticas e as lenticulares ($0 \leq T \leq 1$), e como *late-type* as espirais e irregulares tipo

Magalhães ($T \geq 1.5$). Analisamos também as correlações dos índices entre si e com o tipo morfológico.

3.1 Correlações entre os índices

Todos os parâmetros morfológicos foram desenvolvidos para avaliar um aspecto da morfologia galáctica. Isso implica que os índices não são estatisticamente independentes, às vezes mesmo fortemente correlacionados. A dependência linear entre duas variáveis X e Y pode ser avaliada de maneira simples pelo coeficiente de correlação de Pearson:

$$r(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}. \quad (3.1)$$

A covariância σ_{XY} contém a informação sobre a correlação e os desvios padrão σ agem como uma normalização. O desvio padrão e, em menor grau, as covariâncias são altamente afetados pelos pontos divergentes (*outliers*). Devido aos precedimentos inteiramente automatizados, nossos resultados podem apresentar tais pontos.

Para obter um indicador de correlação mais robusto, pouco ou não sensível aos pontos divergentes, usamos uma pseudocovariância $\overline{\sigma}_{XY}$ que definimos com medianas ao invés de médias (veja Apêndice A). Consequentemente, o desvio padrão é substituído pela mediana dos desvios absolutos à mediana, notada $\overline{\sigma}$. O coeficiente de Pearson da Equação 3.1 se torna:

$$r(X, Y) = \frac{\overline{\sigma}_{XY}}{\overline{\sigma}_X \overline{\sigma}_Y}. \quad (3.2)$$

Dessa forma, os pontos muito afastados da mediana, mesmo que de valores completamente aberrantes, não afetam o cálculo do coeficiente de Pearson. Eventuais pontos divergentes, que poderiam quebrar a dependência das duas variáveis, são negligenciados e, mesmo assim, a pseudocovariância permite medir a correlação. Mas, por outro lado, com dados corretos e precisos, as medianas reduzem o peso dos pontos distantes, o que tem como consequência subestimar as correlações fracas.

3.2 Índices comuns

Em nosso trabalho, concentramo-nos nos seis índices morfológicos mais comumente encontrados na literatura, que são os seguintes, em ordem cronológica de criação:

- O parâmetro de Sérsic n (Sérsic 1963);
- O índice de concentração C (Bershady et al. 2000; Kent 1985);
- O índice de assimetria A (Schade et al. 1995);
- O índice de *clumpiness* S (Conselice 2003; Takamiya 1999);
- O coeficiente de Gini G (Abraham et al. 2003; Lotz et al. 2004);
- O índice M_{20} (Lotz et al. 2004).

Alguns autores fazem uma diferença estrita entre as medidas paramétricas, resultantes de um ajuste a uma lei definida *a priori*, e as medidas não-paramétricas, que não assumem qualquer modelo subjacente. O termo “índice” é, conforme essa modalidade, reservado para medidas não-paramétricas. Isso explica por que se fala do “parâmetro” de Sérsic, que provém do ajuste da lei de Sérsic. Não obstante, se encontra amiúde a expressão índice de Sérsic. Já o coeficiente de Gini foi emprestado do mundo socioeconômico, e a sua denominação inicial foi conservada.

3.2.1 O parâmetro de Sérsic

A distribuição radial do brilho superficial de uma galáxia é uma propriedade chave do ponto de vista da morfologia. A noção de distribuição radial é baseada na aproximação de que o brilho superficial projetado – ou seja, a própria imagem de uma galáxia – depende somente da distância ao centro galáctico e não do azimute. Essa é a hipótese da simetria rotacional. Assim, uma galáxia descrita por uma função bidimensional de intensidade $I(x, y)$ é expressa com a ajuda de uma única variável r :

$$I(x, y) \equiv I(r, \theta) = I(r, \bar{\theta}) = I(r). \quad (3.3)$$

A intensidade radial $I(r)$ é frequentemente chamada de perfil de luz. Ela é medida construindo a curva de crescimento.

A função $I(r)$ traz informações misturadas sobre a geometria da galáxia e sobre a densidade estelar. Ora, essas duas propriedades são radicalmente diferentes para uma

galáxia elíptica a uma galáxia espiral. Isso explica por que o perfil de luz é uma ferramenta fundamental para medir a morfologia. A primeira pessoa a haver formalizado a distribuição radial do brilho superficial foi de Vaucouleurs (de Vaucouleurs 1948, 1958), que doou seu nome a uma função que ajusta bem o perfil de luz das galáxias elípticas, assim como o dos bojos. Mais tarde, Sérsic (1963) generalizou o perfil de de Vaucouleurs deixando um parâmetro livre, o próprio índice de Sérsic, notado n . A lei de Sérsic pode ser escrita assim:

$$I(r) = I_0 \exp(-k \cdot r^{1/n}). \quad (3.4)$$

O parâmetro I_0 é a intensidade no centro galáctico do modelo, e k é um terceiro parâmetro do ajuste, generalizando a função. Encontra-se com frequência uma expressão ligeiramente diferente para a lei de Sérsic, com um raio efetivo r_e que engloba a metade do fluxo total do modelo, e um fator I_e que é a intensidade em r_e .

Grandes valores de n descrevem um perfil muito concentrado no centro, mas simultaneamente muito elevado na parte externa. Pequenos valores de n descrevem um perfil pouco concentrado, mas que tende muito rapidamente a 0. Veja, por exemplo, Graham & Driver (2005) para um resumo conciso sobre o perfil de luz e a lei de Sérsic.

Concretamente, para medir o parâmetro de Sérsic, constrói-se, ponto por ponto, o perfil radial da galáxia. Ou seja, calcula-se a intensidade média em anéis elípticos cujos eixos aumentam de um pixel até o raio total da galáxia. Esse processo é ilustrado na segunda coluna da Figura 3.1, e os perfis obtidos são representados nos gráficos da terceira coluna (linha preta). O perfil da galáxia espiral, na segunda linha, mostra claramente dois comportamentos diferentes que correspondem a duas estruturas distintas: o bojo e o disco.

O ajuste da função de Sérsic é realizado sobre o logaritmo do perfil, a fim de não dar um peso importante demais às intensidades elevadas, na zona mais interna da galáxia. Para descartar um eventual bojo, que tem um perfil diferente do disco, o centro galáctico não é considerado. O ajuste é feito pelo método habitual dos mínimos quadrados entre $0.2r_{Petro}$ e $1r_{Petro}$. A função ajustada, bem como a sua equação, são indicadas em vermelho nos gráficos da Figura 3.1.

A lei de de Vaucouleurs é um caso particular da lei de Sérsic com um parâmetro $n=4$. Ela descreve bem o perfil das galáxias elípticas, assim como os bojos. Um valor

3. MORFOLOGIA QUANTITATIVA

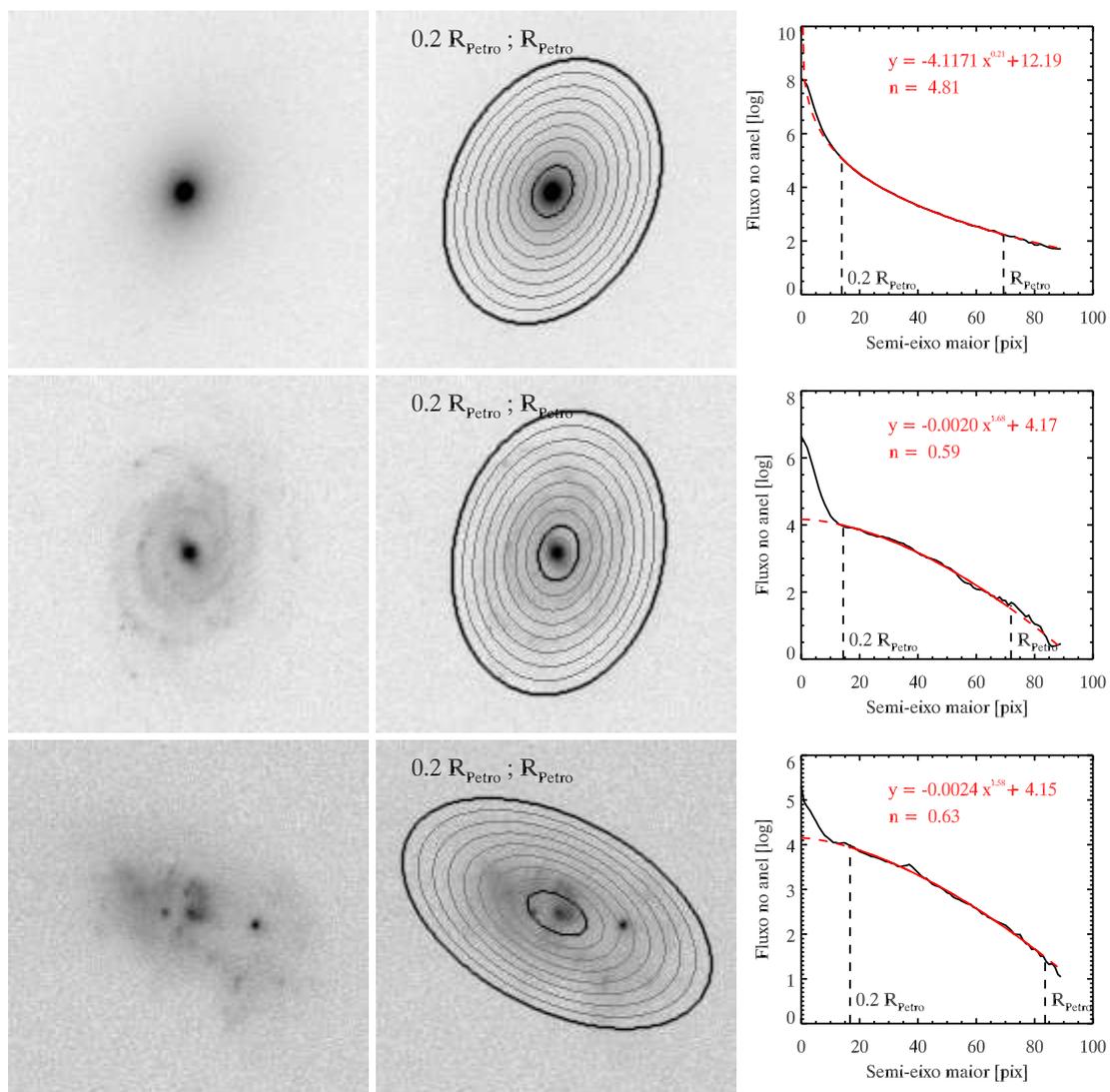


Figura 3.1: Processo de computação do índice de Sérsic n ; três exemplos de uma galáxia elíptica (*Fukugita1849*), uma espiral (*Fukugita0443*) e uma irregular (*Fukugita0022*).

de $n=1$ define o perfil exponencial observado para os discos. Espera-se, portanto, que as galáxias *late-type* sejam descritas por valores de n ao redor de 1, e as *early-type* por valores maiores, ao redor de 4.

A Figura 3.2 esquerda mostra a distribuição do parâmetro de Sérsic para o catálogo morfológico inteiro (linha preta), e para as galáxias *early-type* (vermelho) e *late-type* (azul). Embora se espalhe em um amplo intervalo de 0.2 a 100, o parâmetro de Sérsic toma os valores esperados por tipo morfológico. A separação acontece em torno de $n=2$; no entanto, é longe de ser eficiente, pois há uma fração não negligenciável de galáxias *early-type* na região das *late-types* e vice-versa. Isso mostra que um único índice morfológico nunca basta para reproduzir fielmente uma classificação.

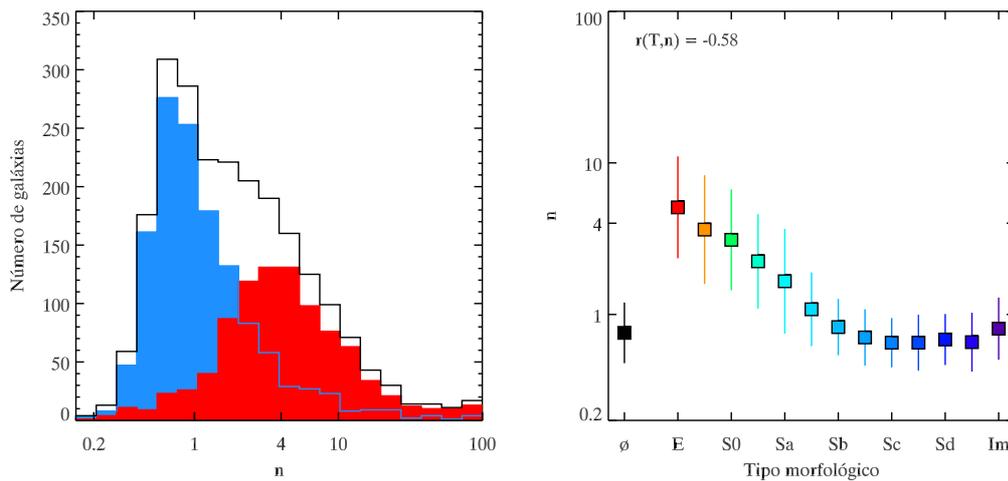


Figura 3.2: Esquerda: distribuição de n . A linha preta se refere ao catálogo morfológico inteiro, o histograma vermelho às galáxias *early-type* e o azul às galáxias *late-type*. Direita: valor de n em função do tipo morfológico. Os quadrados representam a mediana para cada tipo e as barras indicam a dispersão de um desvio absoluto mediano.

Na direita da Figura 3.2, o gráfico mostra o comportamento do parâmetro de Sérsic em função do tipo morfológico. Os pontos são as medianas de todos os valores dentro de um tipo dado, e as barras indicam as dispersões pelo desvio absoluto mediano ($1\bar{\sigma}$). O parâmetro de Sérsic acompanha a sequência de Hubble até o tipo Sc. Mas a partir daí, n fica inalterado, e até aumenta novamente um pouco para as irregulares. As barras de dispersão revelam que dentro de cada tipo a nuvem de pontos se distribui por regiões

3. MORFOLOGIA QUANTITATIVA

ocupadas pelas nuvens de vários outros tipos. Como indicado no gráfico, o coeficiente de correlação entre o parâmetro de Sérsic e o tipo morfológico é $r(T, n) = -0.58$.

O parâmetro de Sérsic, como todos os índices do tipo concentração (n , C e G), é relacionado com a massa ou a luminosidade (Blanton et al. 2003; Zamojski et al. 2007). Em um contexto mais específico, Graham & Driver (2007) propõe até uma relação entre o parâmetro de Sérsic do bojo de galáxias espirais e a massa do buraco negro central supermassivo.

A distribuição espacial da componente fria de uma galáxia é um aspeto tão importante da morfologia que surgiu um método de análise exclusivamente dedicado a isso: a decomposição em “*sersiclets*” (Andrae et al. 2011b; Ngan et al. 2009). É um método híbrido entre um ajuste bidimensional do perfil de Sérsic e uma decomposição da imagem em “*shapelets*” (Refregier 2003), uma variante dos polinômios de Hermite.

3.2.2 O índice de concentração

O perfil radial se tornou rapidamente uma ferramenta indispensável para quantificar a morfologia. Mas medir uma propriedade por um ajuste requer assumir um modelo arbitrário – ou, pelo menos, empírico. Para escapar dessa restrição, a maioria dos estudos morfológicos recentes prefere usar o índice de concentração que o parâmetro de Sérsic.

Basicamente, não há grande diferença entre essas duas grandezas. Nos dois casos, o propósito é medir a declividade do perfil radial. O índice de concentração C cumpre essa tarefa calculando em qual raio se concentra uma certa fração do fluxo galáctico. O fluxo em função do raio, ou, em nosso caso, em função do semieixo maior, é a curva de crescimento.

Quase similar ao perfil radial, a curva de crescimento é construída somando o fluxo da galáxia dentro de elipses, cuja abertura aumenta. A curva de crescimento é assim a integral, em coordenadas polares, do perfil radial. Na Figura 3.3, as curvas de crescimento das nossas três galáxias espécimes são reproduzidas na última coluna. O fluxo é expresso em porcentagem do fluxo total, definido como o fluxo dentro do raio de Petrosian r_{Petro} . Determina-se R_{20} como o raio correspondendo a 20% do fluxo; da mesma maneira, determina-se R_{80} . As elipses dos semieixos R_{20} e R_{80} são desenhadas em negrito na coluna do meio.

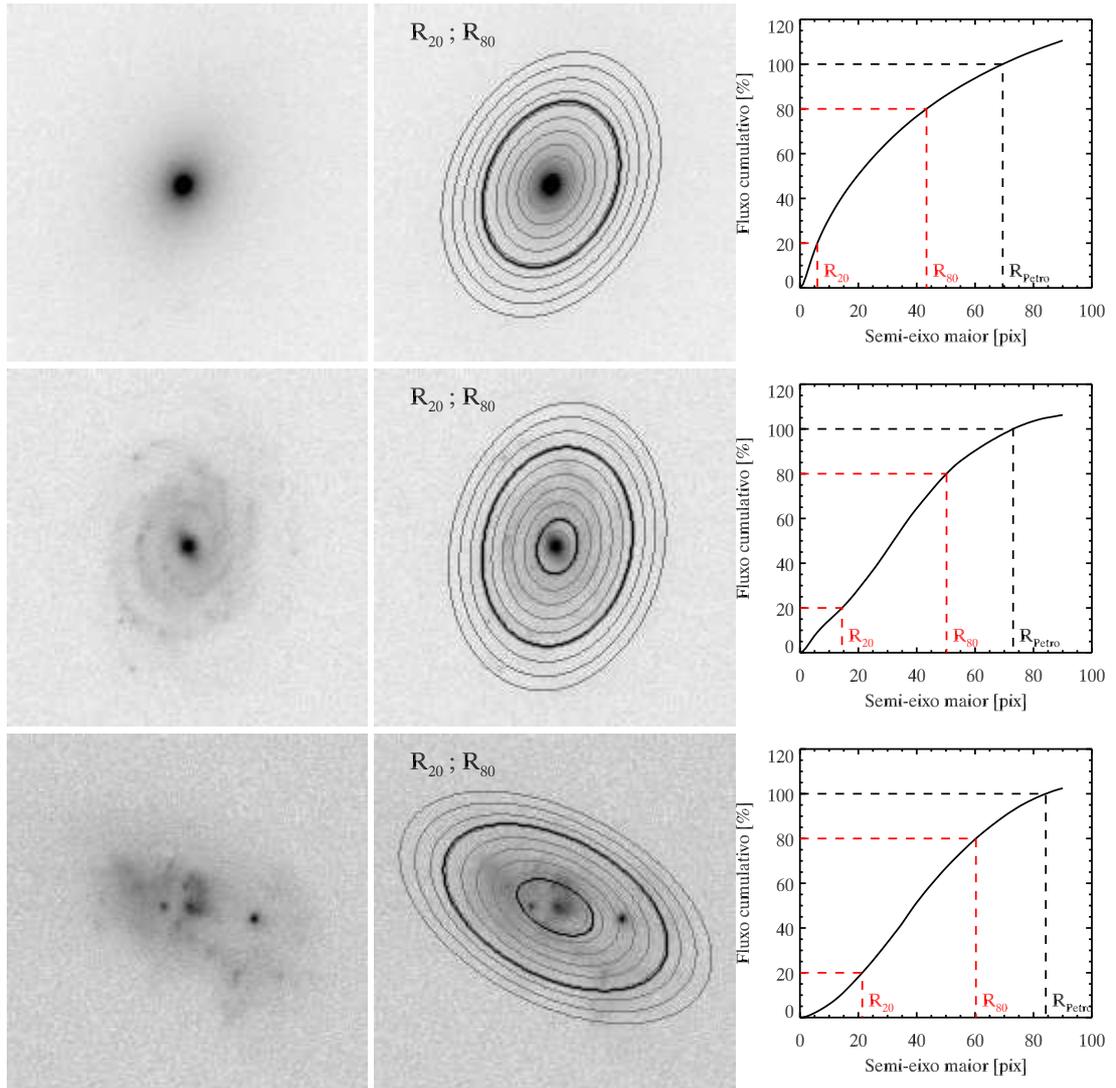


Figura 3.3: Processo de computação do índice de concentração C ; três exemplos sobre *Fukugita1849*, *Fukugita0443* e *Fukugita0022*.

3. MORFOLOGIA QUANTITATIVA

O índice de concentração pode agora ser definido como a razão R_{20}/R_{80} . Devido à grande amplitude dos valores atingidos, C é geralmente avaliado em logaritmo e multiplicado pelo fator 5, normas que aplicamos em nosso trabalho: $C = 5 \log(R_{20}/R_{80})$.

Na Figura 3.4, à direita, a distribuição de C para a amostra toda já sugere uma bimodalidade pelo pico duplo da linha preta. Separando em *early*- e *late-types*, duas curvas de sino aparecem bem distintamente, provando a capacidade de C para delimitar classes. No entanto, a interpenetração dos dois sinos destaca uma zona de dúvida ao redor de $C=3$. A incerteza nessa região pode ser real (existem galáxias *late-type* com C maior do que 3.5) ou artificial (problema durante a redução dos dados ou erro estatístico no cálculo dos índices). Tais zonas de incerteza, nos limites dos tipos morfológicos, são naturais e se encontram para todos os índices. As classificações automáticas têm justamente o papel de escolher um método para levar ao máximo as incertezas. Aumentando a dimensão do espaço dos parâmetros, uma combinação apropriada de índices complementários pode reduzir significativamente as zonas de dúvidas.

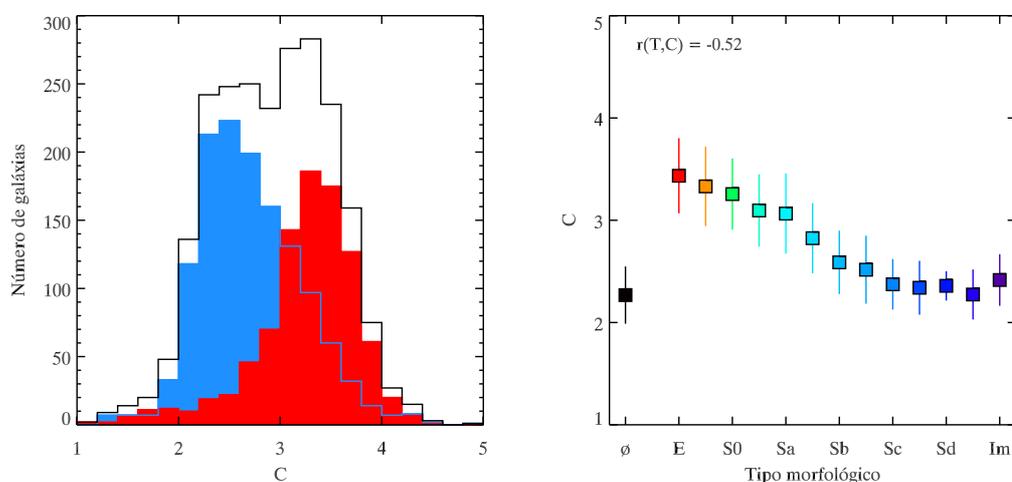


Figura 3.4: Distribuição de C (esquerda) e mediana de C em função do tipo morfológico (direita).

No lado direito da Figura 3.4, C em função do tipo mostra um comportamento muito similar àquele de n . Entre os tipos E e Sc, em média, C é proporcional a T . Mas a partir de Sc, C fica quase constante. Na perspectiva de uma classificação automática, C seria então eficiente entre E e Sc, mas deveria ser combinado ao menos com um

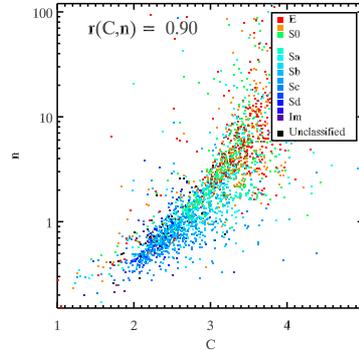


Figura 3.5: Relação entre C e n . A cor dos pontos indica o tipo morfológico tal como determinado por Fukugita et al. (2007).

outro índice que mostra uma dependência do tipo entre Sc e Im . Essa constatação é a mesma para todos os índices morfológicos do tipo concentração.

Como esperado, C é bastante correlacionado com o parâmetro de Sérsic n (Figura 3.5). Os dois índices avaliam a mesma propriedade, a saber, a distribuição radial da luz. Notem que n é representado em escala logarítmica¹, e que o coeficiente de correlação é calculado para $\log(n)$, sem o qual seria muito menor.

C e n são dois indicadores morfológicos irmãos e têm características semelhantes. O uso de ambos na mesma classificação seria redundante e não traria melhoria.

3.2.3 O índice de assimetria

O índice de assimetria rotacional avalia se a galáxia mostra as mesmas estruturas nos dois lados do centro. A ideia é então comparar a imagem da galáxia com a sua imagem meia-volta girada.

A imagem original é chamada de I . Através de uma rotação de 180° de I , obtemos a imagem I_{180} . Essas duas imagens são subtraídas uma da outra para formar a imagem residual $I - I_{180}$, como representada na coluna do meio da Figura 3.6. No resíduo, os elementos de estrutura que possuem uma contraparte do outro lado do ponto de rotação desaparecem. Só sobrevivem as estruturas assimétricas. Como o resíduo é uma subtração de duas vezes a mesma imagem, o seu fluxo total é nulo. Para medir a quantidade de fluxo presente nas estruturas assimétricas, é portanto necessário tomar o

¹O logaritmo já aparece explicitamente na definição de C , mas não naquela de n .

3. MORFOLOGIA QUANTITATIVA

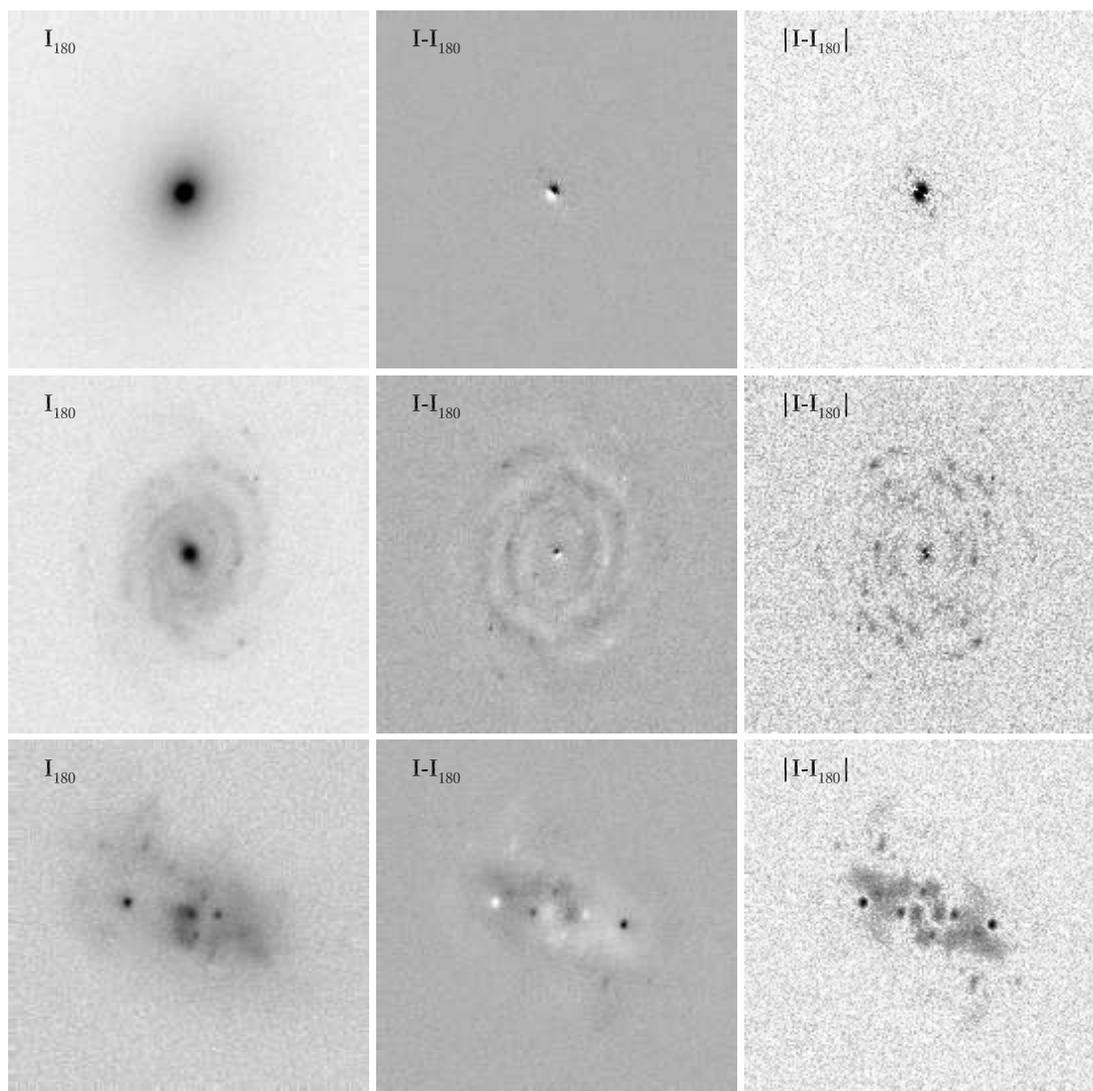


Figura 3.6: Processo de computação do índice de assimetria A .

valor absoluto do resíduo². Ainda se precisa normalizar o fluxo assimétrico utilizando o fluxo da galáxia. A assimetria A se mede pelo quociente dos fluxos no resíduo e na imagem: $A = |I - I_{180}| / I$.

Resumindo, A é definido como a fração do fluxo galáctico provindo de estruturas que não têm gêmeo simétrico. Ele é um número entre 0 (completamente simétrico) e 1 (completamente assimétrico).

O centro de rotação é o ponto que minimiza A , ou seja, não é sistematicamente o centro de luz. Um procedimento iterativo computa vários A para centros de rotação em uma sequência de grades de pontos. A primeira grade usa o centro habitual. Em seguida, ela é deslocada e refinada até achar o valor mínimo.

O ruído do céu tem um efeito considerável no cálculo de A . Os píxeis são aleatórios no resíduo e podem participar de maneira significativa do seu fluxo. Para uma medida rigorosa, a contribuição do ruído deve ser retirada do valor calculado sobre a imagem. Abordamos esse problema de maneira analítica, calculando qual é, no resíduo, o fluxo médio de uma imagem de píxeis para os quais a distribuição é normal (Veja Capítulo 5, Equação 3).

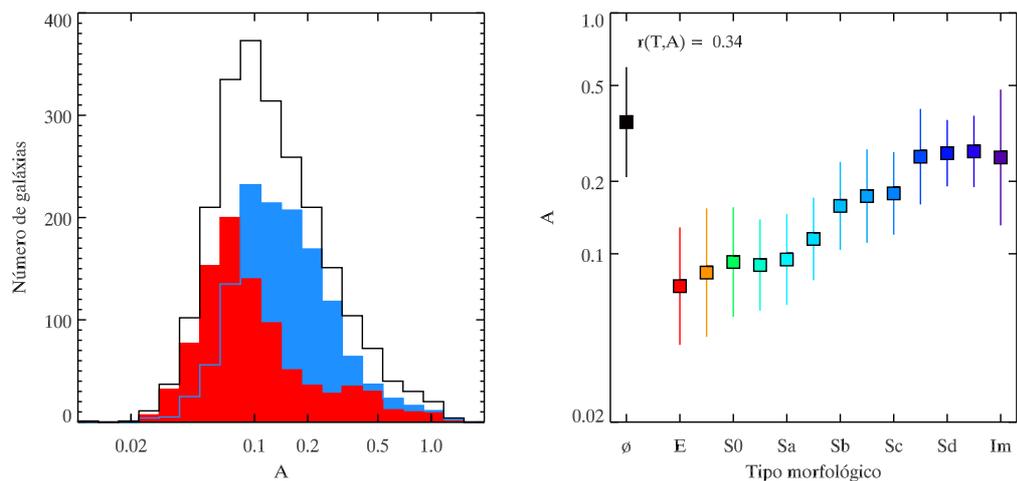


Figura 3.7: Distribuição de A (esquerda) e mediana de A em função do tipo morfológico (direita).

²Às vezes, encontra-se o quadrado do resíduo no lugar do valor absoluto.

3. MORFOLOGIA QUANTITATIVA

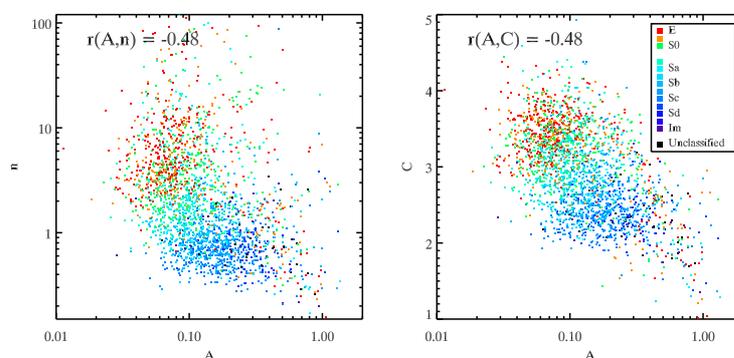


Figura 3.8: Relações entre A e os índices precedentes.

Por muito tempo, o índice de assimetria foi o principal – senão o único – indicador de galáxias interagentes. Além disso, a assimetria é um bom traçador de tipo morfológico. É óbvio que as galáxias elípticas são muito simétricas, enquanto as galáxias irregulares (Im, ou mesmo Sd) contêm picos luminosos distribuídos de maneira disparatada.

Mas entre esses dois extremos, encontramos um contínuo de graus de assimetria. As galáxias Sa têm braços espirais regulares e pouco desenhados sobre o disco, de tal maneira que são bastante simétricas. As Sc têm braços espirais muito visíveis, nos quais as regiões de formação estelar são distribuídas aleatoriamente. Se pode esperar, então, que o índice A aumente ao longo de toda a sequência de Hubble. A Figura 3.7, à direita, mostra que não é exatamente o caso, pois os quatro últimos pontos têm quase o mesmo valor de A . Com a definição e procedimentos escolhidos, as galáxias Im não aparecem mais assimétricas que as Sd.

Na Figura 3.7, constatamos ainda que S não é fortemente correlacionado com o tipo T , o que pode ser visto no índice de correlação que vale apenas 0.34 e nas barras de dispersão amplas. Isso explica por que S consegue só modestamente separar as *early-types* das *late-types*.

Na Figura 3.8, o índice A é comparado com os índices precedentes n e C . Em ambos os planos existem correlações, mas não são tão fortes, o que ilustra que são medidas diferentes, no entanto complementares. Nos planos A - n e A - C , as nuvens de pontos são adequadamente espalhadas e determinam regiões dominadas por tipos diferentes. Em vista disso, julga-se ser essencial juntar A a um índice do tipo concentração

em uma classificação automática.

3.2.4 O índice de *clumpiness*

A palavra *clumpiness* é de difícil tradução em português, por isso é empregada normalmente em inglês. A *Clumpiness* é a propriedade de ser *clumpy*, ou seja, formado de *clumps*. Um *clump* é uma massa irregular, um amontoado, um tufo, palavra frequentemente usada para árvores ou plantas. Em outros contextos, encontra-se, às vezes, o neologismo “grumosidade” que poderia ser usado como tradução de *clumpiness*.

O índice de *clumpiness*, nomeado S , procura medir a fração do fluxo da galáxia proveniente dos *clumps* de luz. Para essa finalidade, e do mesmo modo que para o índice A , cria-se uma imagem residual. Idealmente, o resíduo contém apenas as estruturas de pequenas escalas, os próprios *clumps*. Ele é o resultado da subtração da imagem original por uma imagem suavizada.

A suavização (*smoothing*, em inglês) é realizada por uma convolução com uma caixa quadrada³ com lado de $1/4r_{Petro}$. Na imagem I_{suave} criada desse modo, os elementos de pequeno tamanho são dissolvidos, e apenas os elementos maiores do que o filtro são visíveis. Ao contrário, no resíduo $I - I_{suave}$, apenas os pequenos elementos aparecem, o que permite definir os *clumps*. A imagem filtrada e o resíduo são representados nas duas primeiras colunas da Figura 3.9.

O centro galáctico não é considerado como um *clump*. Antes de detectar os *clumps*, o centro da galáxia é escondido por uma máscara com a mesma forma que a isofota de Petrosian, mas com dimensões quatro vezes menores. Os contornos dessa máscara são desenhados acima da imagem do resíduo.

Agora, os *clumps* são identificados por um *threshold* a 3σ do ruído. As estruturas detectadas são mostradas na última coluna da Figura 3.9. O índice de *clumpiness* é medido pela razão entre o fluxo dos *clumps* e o fluxo da imagem: $S = (I - I_{suave})_{3\sigma} / I$.

Numerosas galáxias *early-type* têm um S estranhamente elevado, como pode ser visto no histograma da Figura 3.10. São galáxias com uma estrela do campo no primeiro plano ou, com mais frequência, com uma galáxia tão vizinha que fica dentro da área definida pela isofota de Petrosian. Apesar disso, S é, em média, um indicador

³Parodi & Binggeli (2003) usam um filtro mediano.

3. MORFOLOGIA QUANTITATIVA

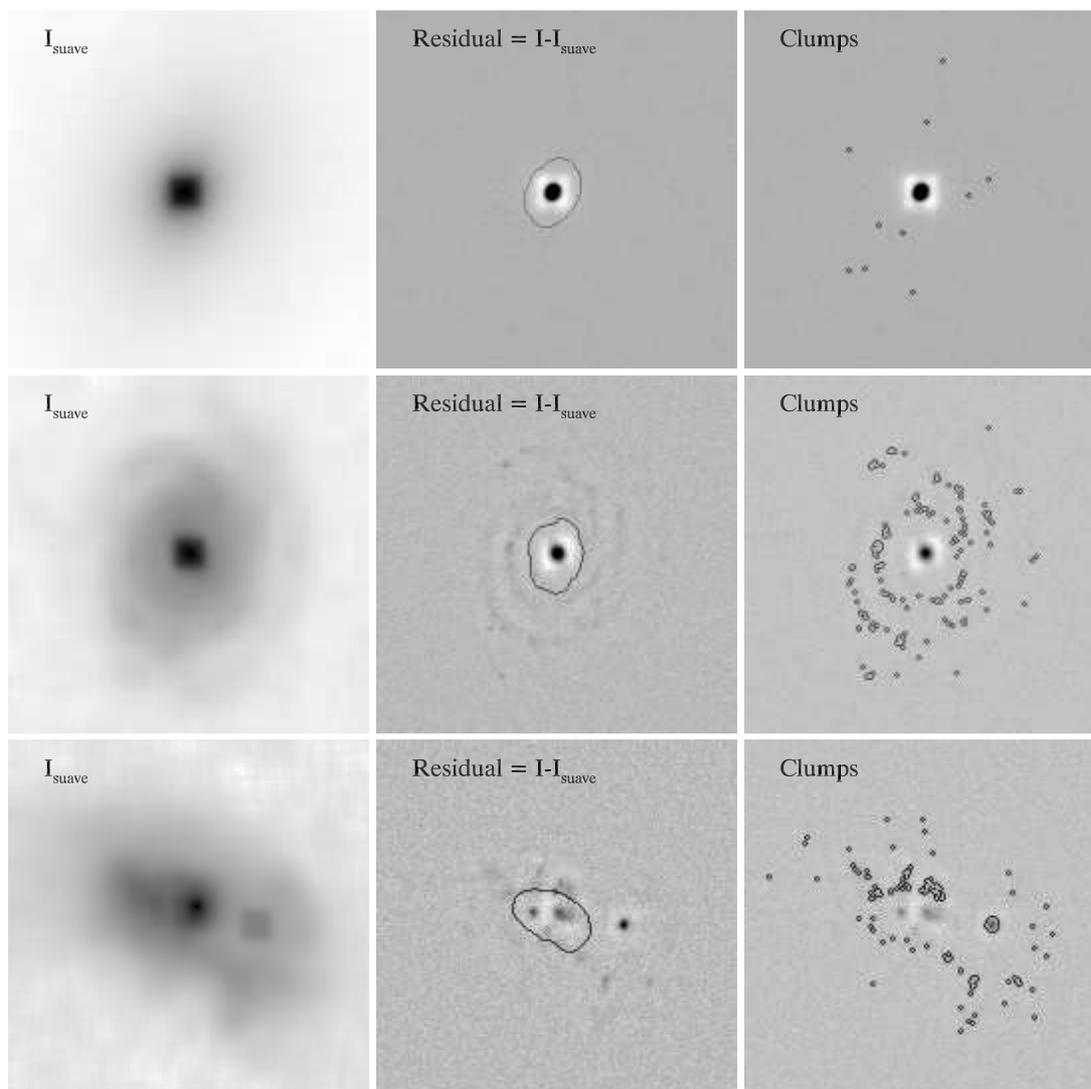


Figura 3.9: Processo de computação do índice de clumpiness S .

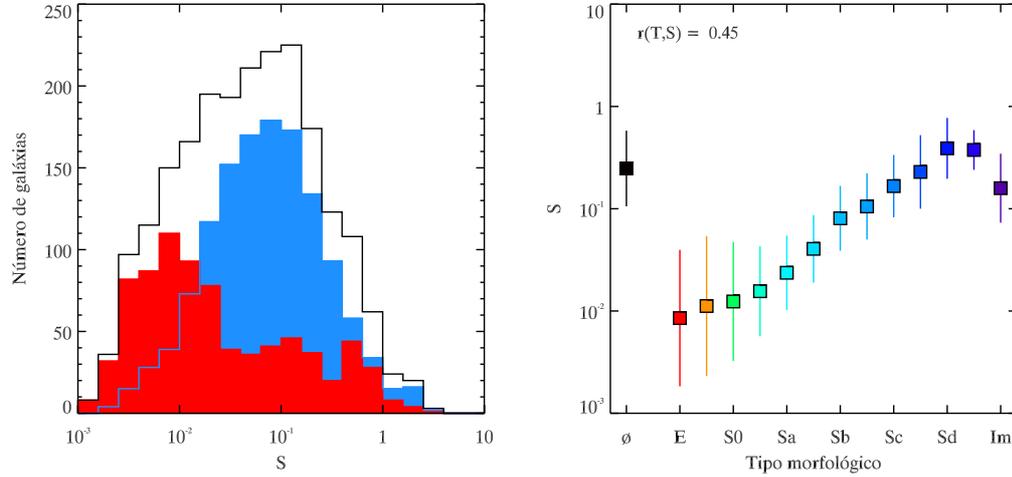


Figura 3.10: Distribuição de S (esquerda) e mediana de S em função do tipo morfológico (direita).

confiável da morfologia entre os tipos E e Sd. Surpreendentemente, são as galáxias Sd que têm, em média, a *clumpiness* máxima; as Im têm um valor de S igual às Sc.

Comparando S com os índices anteriores (Figura 3.11), notamos uma interdependência forte entre S e A . Os *clumps* são picos luminosos que correspondem, sobretudo, a regiões de alta formação estelar. Embora um braço espiral tenha normalmente o seu dobro do outro lado da galáxia, as regiões de alta formação estelar são distribuídas de modo irregular nos braços. Por conseguinte, os elementos assimétricos entrando no cálculo de A são principalmente os próprios *clumps*, e os residuais construídos para A

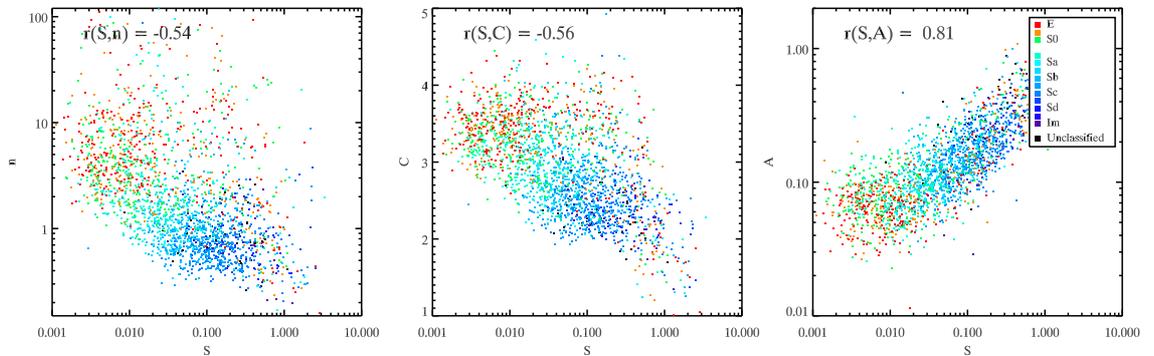


Figura 3.11: Relações entre S e os índices precedentes.

3. MORFOLOGIA QUANTITATIVA

e S são semelhantes. Como os dois índices são definidos da mesma maneira, os seus valores são muito próximos.

Nos planos $S-n$ ou $S-C$, os pontos fora da correlação principal são precisamente as galáxias *early-type* problemáticas do histograma. Efetivamente, os pontos acima da diagonal descendente são galáxias muito concentradas e, ao mesmo tempo, com um índice S alto; são galáxias *early-type* (pontos vermelhos, laranja e verdes). Esses pontos divergentes resultam de considerações metodológicas e não são oportunos no estudo morfológico. Em uma aprendizagem automática, tais pontos não devem entrar na amostra das instâncias de treinamento. Podem ser eliminados por um processo estatístico como desenvolvido no Apêndice A.

Os índices de concentração C , de assimetria A e de *clumpiness* S formam juntos o que se encontra na literatura como a classificação CAS.

3.2.5 O coeficiente de Gini

O coeficiente de Gini é um indicador socioeconômico da desigualdade na distribuição da riqueza. Aplicando a medida sobre os píxeis de uma galáxia, Abraham et al. (2003) adaptaram o coeficiente de Gini para usá-lo como índice morfológico. O Apêndice B descreve como quantificar as desigualdades e detalha, com exemplos, o cálculo do coeficiente de Gini.

Embora o índice de concentração não resulte do ajuste a um modelo – e é, nessa visão, uma generalização do parâmetro de Sérsic –, ele assume, ainda assim, a existência de um centro e de uma certa geometria. O coeficiente de Gini foi introduzido como uma alternativa ao índice de concentração, a fim de resolver o problema do centro e sair da hipótese da simetria rotacional. Assim, ele é mais adaptado às galáxias irregulares.

O coeficiente de Gini é baseado na curva de Lorenz, que é uma representação da distribuição de uma variável. Mais exatamente, a curva de Lorenz indica a fração do total da variável em função da fração do número de eventos⁴.

Para não ser contaminado pelo ruído, o coeficiente de Gini G deve ser calculado com uma abertura pequena. A primeira coluna da Figura 3.12 mostra a isofota de Petrosian (linha externa) e uma segunda abertura menor, da mesma forma mas reduzida

⁴No sentido probabilístico, um evento é o conjunto de resultados possíveis de uma variável.

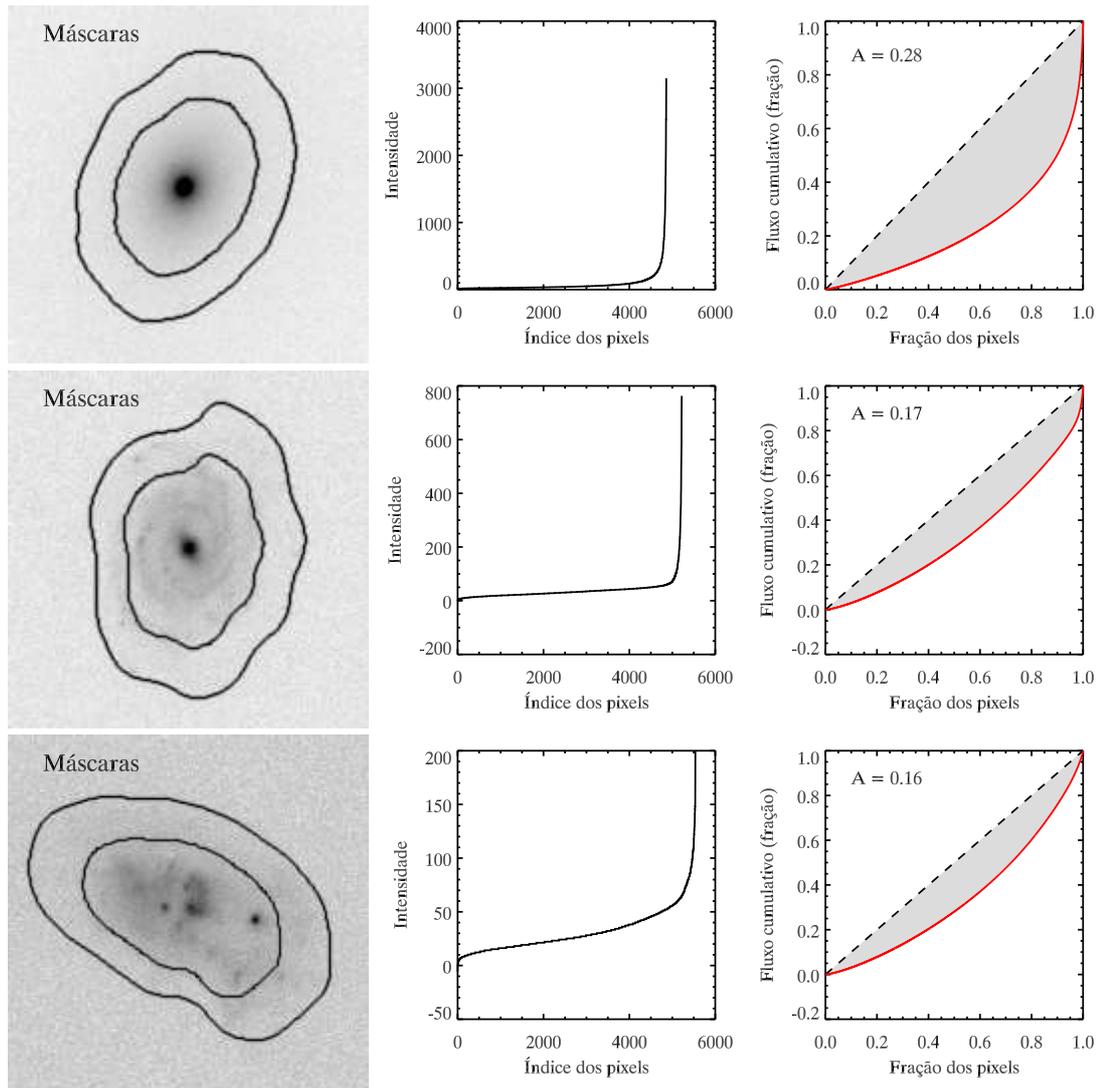


Figura 3.12: Processo de computação do coeficiente de Gini G .

3. MORFOLOGIA QUANTITATIVA

por um terço (linha interna), usada para a computação de G . Os píxeis no interior da área diminuída são ordenados da menor intensidade para a maior, formando a curva da segunda coluna. Após a normalização do número dos píxeis e do fluxo, a integral dessa função, computada como a soma acumulada, define a função de Lorenz (terceira coluna, linha vermelha). A área cinza entre a diagonal e a curva de Lorenz determina a desigualdade: quanto maior a área for, maiores as desigualdades entre as intensidades de píxeis. Com isso, o coeficiente G se define facilmente como o dobro da área cinza (indicada na parte de cima do gráfico).

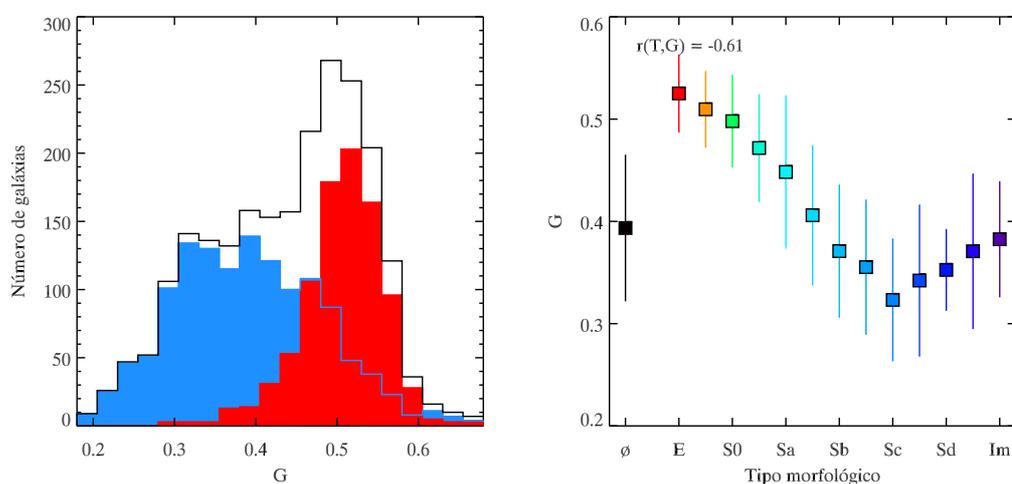


Figura 3.13: Distribuição de G (esquerda) e mediana de G em função do tipo morfológico (direita).

Assim como os outros índices do tipo concentração, G consegue com eficiência separar as galáxias *early-* das *late-type* (Figura 3.14, esquerda). A dependência entre G e o tipo morfológico parece seguir dois regimes lineares mas diferentes (Figura 3.14, direita): a relação é decrescente entre E e Sc, depois crescente entre Sc e Im. O regime da esquerda, decrescente, é dominado pela luz da componente fria, aquela originada nas estruturas de grandes escalas (disco, galáxia como um todo). Nesse regime, a desigualdade na distribuição das intensidades é devida à distribuição radial, ou seja, G é inversamente proporcional a C . No regime da direita, crescente, o perfil radial do disco é o mesmo. Neste regime, a desigualdade das intensidades é devida sobretudo à presença de picos luminosos de pequeno tamanho (*clumps*), ou seja, G é proporcional

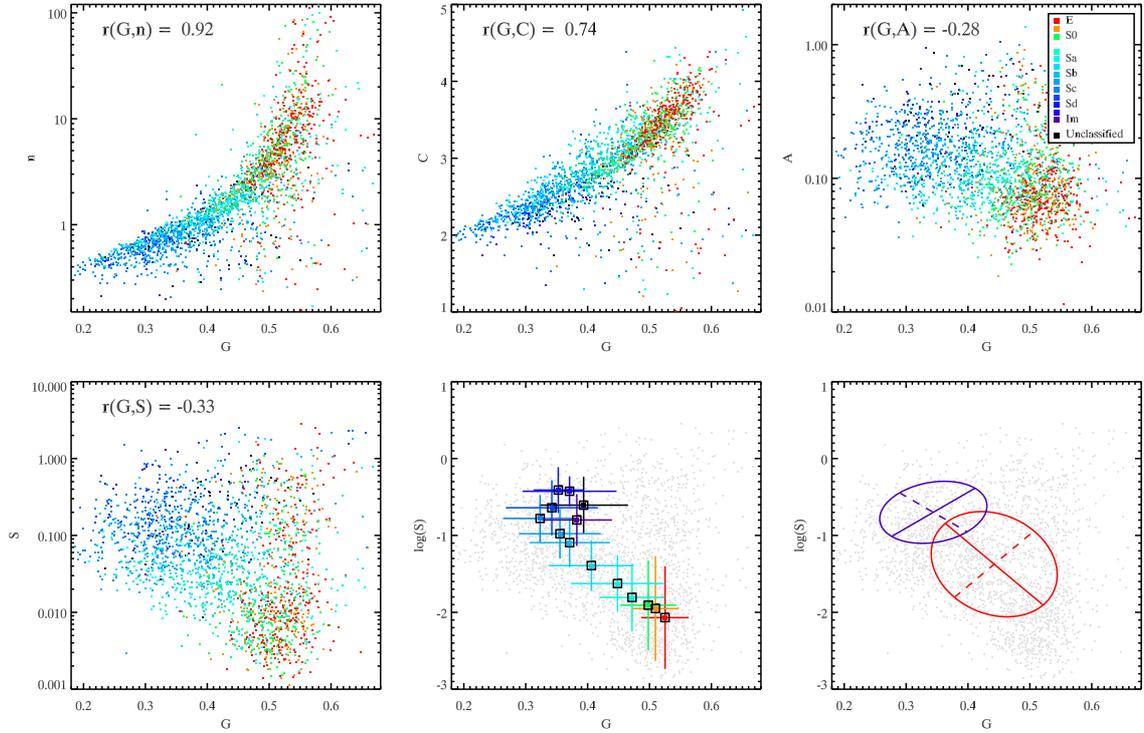


Figura 3.14: Relações entre G e os índices precedentes.

a S . O mínimo de G é alcançado para as Sc , o que significa que as intensidades por pixel dessas galáxias são distribuídas com a maior igualdade. As galáxias Sc são um compromisso entre um perfil radial não muito concentrado e uma quantidade de *clumps* não muito importante.

Na Figura 3.14, os dois regimes de G podem ser vistos nos três planos $G-S$, diferentemente representados na última linha. Com apenas os pontos, é difícil identificar os dois regimes, mas as medianas de cada tipo já sugerem uma relação ao longo da diagonal descendente e uma outra, mais desordenada, no canto superior esquerdo. No último gráfico, as elipses têm como eixo maior (linha contínua) a primeira PC (componente principal) e, como eixo menor (linha tracejada), a segunda PC. A elipse vermelha é obtida para as galáxias do primeiro regime, de tipo E a Sc ; a elipse azul para as galáxias do segundo regime, de tipo Sc a Im.

As fortes correlações entre G e n ou C provam que, como já o interpretamos através da sua definição, o coeficiente de Gini é, em primeiro lugar, um indicador de concen-

3. MORFOLOGIA QUANTITATIVA

tração de luz.

3.2.6 O índice M_{20}

A sigla M_{20} esconde uma expressão bárbara: o momento centrado de segunda ordem dos píxeis mais luminosos contendo 20% do fluxo total da galáxia.

Mostramos na Seção 2.3.2 que os momentos de segunda ordem são uma medida da dispersão de uma variável aleatória e que podem ser usados para descrever uma galáxia como uma gaussiana bidimensional. Aplicando a definição dos segundos momentos sobre a distância dos píxeis ao centro galáctico, em vez da posição vetorial, se obtém o momento centrado do raio. O momento centrado mede a dispersão da distribuição de luz ao redor do centro. Com o índice M_{20} , procuramos saber se os píxeis mais luminosos são localizados perto do centro ou espalhados longe dele.

O índice M_{20} foi introduzido como indicador de galáxias em interação ou, mais precisamente, de galáxias em fusão. Uma das assinaturas de uma recente fusão de galáxias é a presença de um núcleo duplo ou de núcleos múltiplos. Quando os píxeis mais luminosos são repartidos em áreas distintas e afastadas umas das outras, M_{20} se torna muito grande.

Também para as galáxias não interagentes, M_{20} é uma quantidade pertinente para figurar a sequência de Hubble. De fato, nas galáxias *early-type*, os píxeis mais luminosos ficam exclusivamente no centro, confinando M_{20} a valores modestos. Ao contrário, as galáxias na extrema direita da sequência de Hubble são compostas de numerosos picos intensos. Estes participam dos 20% do fluxo mais luminoso, e o momento de segunda ordem é calculado para píxeis distribuídos em diferentes regiões da galáxia, elevando M_{20} a valores importantes.

A primeira etapa para calcular o índice M_{20} é determinar quais são os píxeis mais intensos, que emitem 20% do fluxo total. Os píxeis da galáxia, dentro da isofota de Petrosian, são ordenados do menos luminoso para o mais luminoso. Calcula-se o fluxo cumulativo para constituir a curva de Lorenz, representada na primeira coluna da Figura 3.15. Nessa curva, podemos ler quais píxeis são responsáveis por 20% do fluxo. Note que, no caso da galáxia elíptica da primeira linha, apenas 1% dos píxeis mais intensos contêm 20% do fluxo total da galáxia. Os píxeis são, em seguida, localizados nas galáxias, como mostrado na segunda coluna.

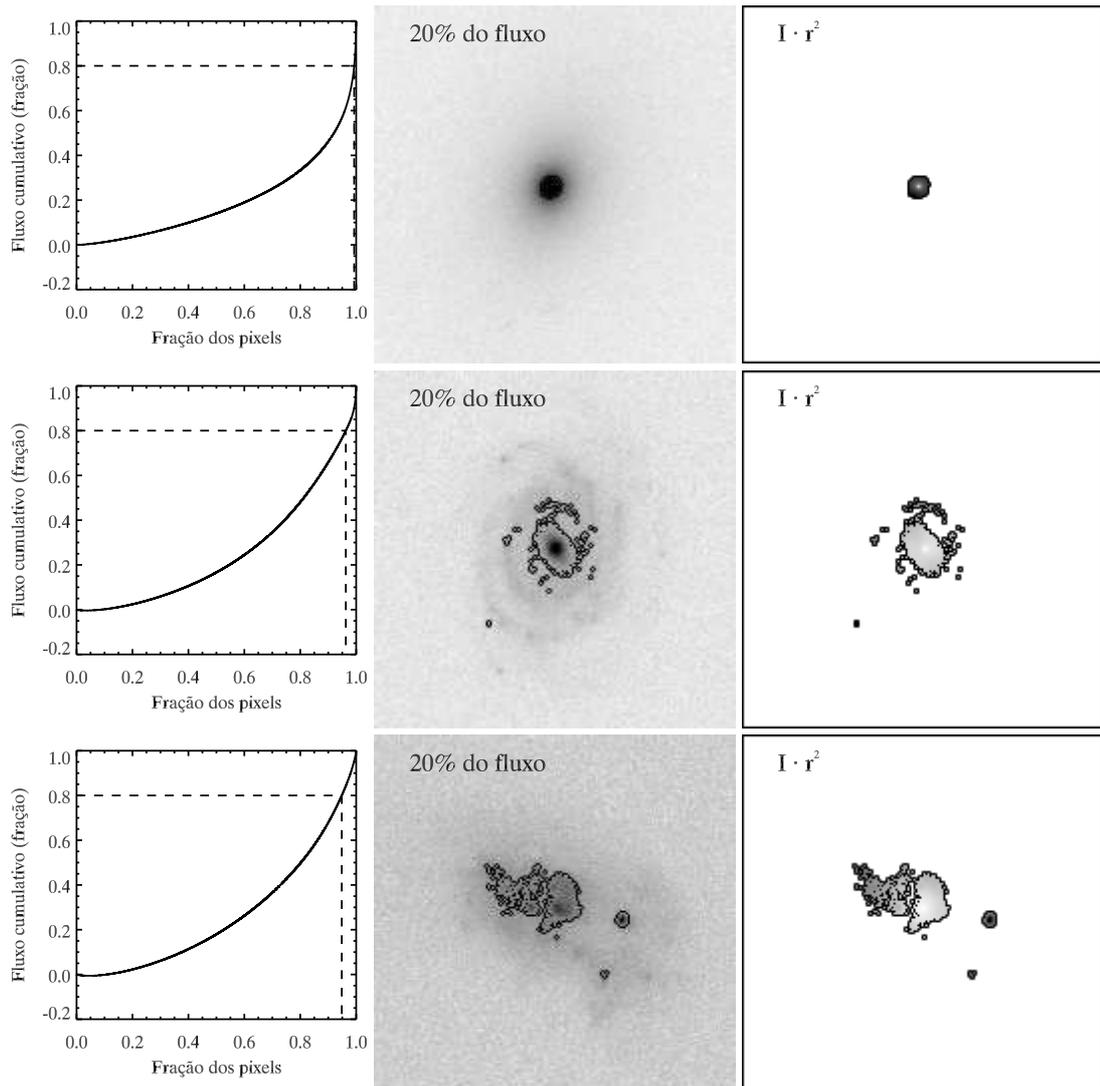


Figura 3.15: Processo de computação do índice M_{20} .

3. MORFOLOGIA QUANTITATIVA

Para computar o momento de segunda ordem, a intensidade dos píxeis supracitados é multiplicada pelo quadrado da sua distância ao centro galáctico. Matematicamente, calculamos a grandeza: Ir^2 . Dessa maneira, os píxeis perto do centro, mesmo com uma intensidade elevada, têm um peso quase nulo, enquanto aqueles remotos, mesmo pouco intensos, adquirem um peso importante. A última coluna da Figura 3.15 representa as regiões da galáxia emitindo 20% do fluxo, nas quais as intensidades foram multiplicadas pelo quadrado da distância do pixel ao centro galáctico.

A quantidade assim obtida não tem muito sentido físico, pois depende do tamanho da imagem. Ela deve ser normalizada pelo momento de segunda ordem da galáxia inteira. Finalmente, o índice M_{20} é computado da seguinte maneira: $M_{20} = \log(\{Ir^2\}_{20}/Ir^2)$.

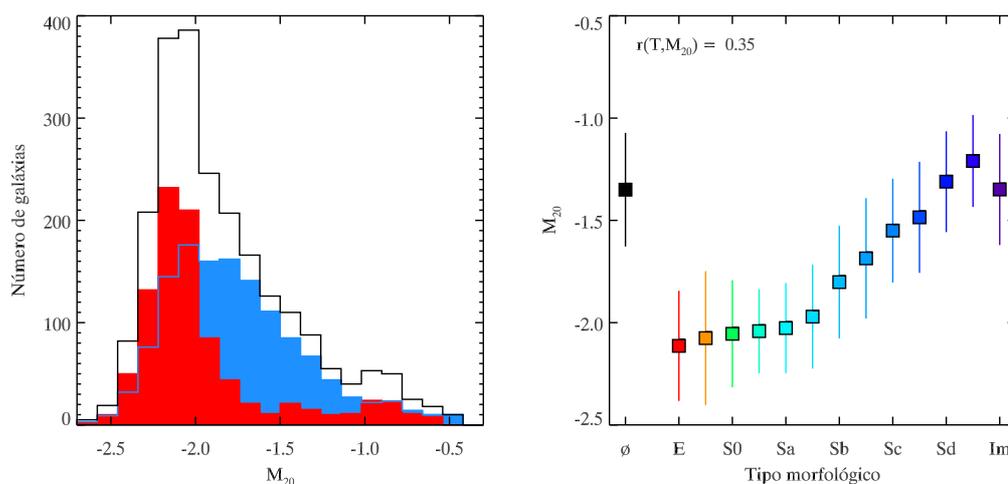


Figura 3.16: Distribuição de M_{20} (esquerda) e mediana de M_{20} em função do tipo morfológico (direita).

M_{20} não consegue separar de maneira ideal as *early-types* e as *late-types*, como ilustra a Figura 3.16, à esquerda. Isso se explica pelo fato de que certos bojos de galáxias discos são muito luminosos e, se os braços espirais não abrigam nenhuma região mais intensa do que o bojo, os píxeis luminosos são todos localizados no centro; tipicamente é o caso para as espirais Sa. O histograma azul indica que há numerosas galáxias *late-type* com valores de M_{20} muito baixos.

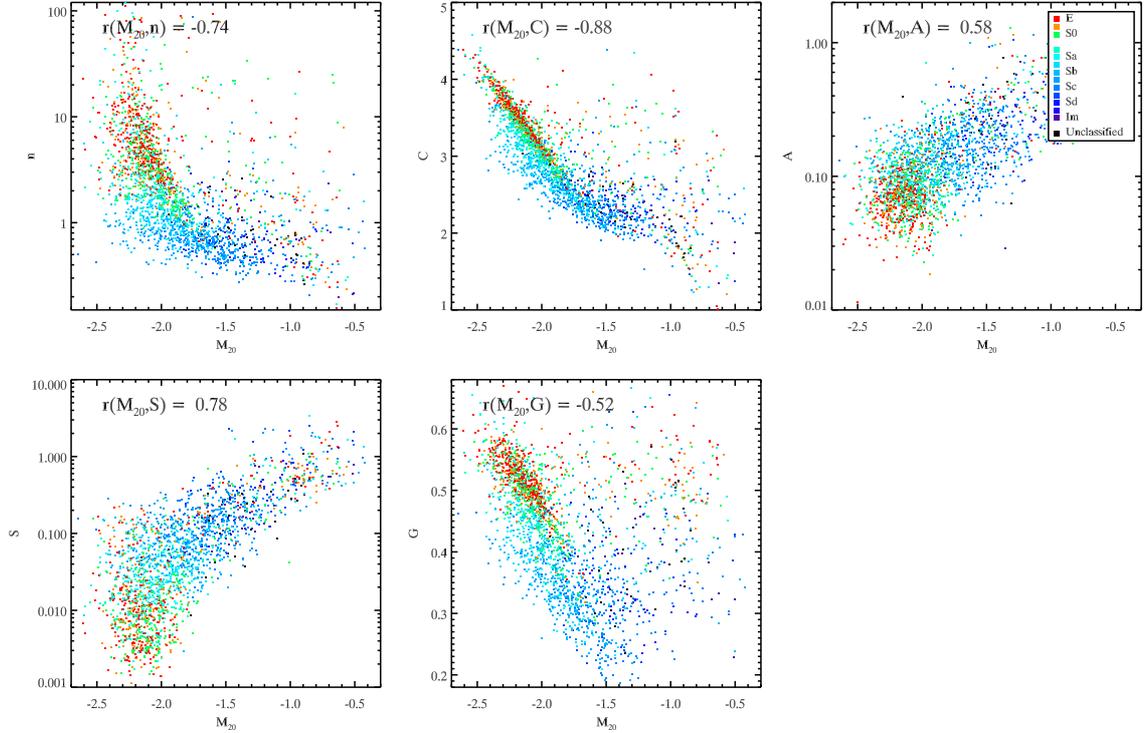


Figura 3.17: Relações entre M_{20} e os índices precedentes.

A dependência entre M_{20} e o tipo morfológico (gráfico da direita, Figura 3.16) exibem claramente dois regimes. Entre E e Sa, M_{20} aumenta pouco ao longo da sequência de Hubble. Nesse regime, em média, todos os píxeis contendo 20% da luz intensa ficam no centro galáctico. Como o momento de segunda ordem mede a dispersão de uma distribuição, M_{20} é aqui proporcional aos índices do tipo concentração. Essa dependência é muito clara no plano M_{20} -C da Figura 3.17, onde os pontos vermelhos, laranja e verdes se agrupam numa reta descendente bastante estreita.

O segundo regime, entre Sb e Sd, corresponde às galáxias nas quais os *clumps* participam do fluxo originado nas zonas mais luminosas. Nesse regime, quanto mais há fluxo nos *clumps*, maior é M_{20} ; o que justifica que M_{20} é proporcional a S somente para as galáxias Sb e seguintes (pontos azuis escuros no plano M_{20} - S).

O plano G - M_{20} é usado por Lotz et al. (2004) como indicador de fusão. As galáxias situadas acima da nuvem de pontos, sobre a diagonal descendente, são consideradas como galáxias em fusão. Veja o Capítulo 5 para maiores detalhes.

3.3 Dois novos índices morfológicos

A fim de orientar a nossa pesquisa para estudos específicos, desenvolvemos dois novos índices morfológicos. O primeiro é o índice de anelipticidade, que mede a deformação da isofota externa da galáxia. O segundo é chamado de índice de granulação e avalia a distribuição radial das regiões de formação estelar. Os dois índices são elaborados com o propósito de serem dedicados particularmente às galáxias irregulares.

A anelipticidade tem como única aplicação identificar as galáxias em interação. Efetivamente, a presença de uma massa vizinha cria perturbações gravitacionais que deformam a galáxia. A parte externa da galáxia é especialmente distorcida por causa de forças de maré. Quando a colisão é mais avançada, as duas galáxias não são o suficiente separadas espacialmente. Se a isofota de Petrosian encerra duas ou mais galáxias, elas são consideradas como em colisão e são estudadas como um só sistema. Tenhamos em mente que galáxias não interagentes podem ser consideradas como tais por efeito de projeção.

A granulação pode ser empregada como índice morfológico para todos os tipos da sequência de Hubble. No Capítulo 5, Seção 6, servimo-nos dela para analisar os lugares de formação estelar durante uma interação.

Os nossos dois novos índices são definidos e examinados em detalhes no Capítulo 5, a anelipticidade na Seção 4.1 e a granulação na Seção 4.2. Nas seções subsequentes, limitamo-nos à mesma abordagem passo a passo que nas anteriores.

3.3.1 O índice de anelipticidade

A isofota de Petrosian foi introduzida com o objetivo de definir meticulosamente o limite externo da galáxia. Podemos agora quantificar a deformação da galáxia usando a figura geométrica da isofota de Petrosian.

O índice de anelipticidade α determina se a isofota de Petrosian tem uma forma regular ou não. Quando a isofota não faz meandro, a área A no seu interior é grande; ou inversamente, o seu perímetro P é pequeno. Para avaliar a eficiência da isofota em maximizar a área para um perímetro fixado, usamos o parâmetro adimensional P^2/A . A elipticidade influencia esse quociente fortemente. Para quebrar a dependência, o quociente é normalizado com ajuda de uma elipse.

A normalização é a quantidade inversa A_{ell}/P_{ell}^2 calculada sobre uma elipse de mesma elipticidade que a isofota. A elipse é determinada a partir dos momentos de segunda ordem da isofota, como na Seção 2.3. Na segunda coluna da Figura 3.18, ela é sobreposta à isofota de Petrosian. Conhecendo seus eixos, conhecemos também a elipticidade. A normalização, que é somente função da elipticidade, pode ser agora avaliada, como ilustrado na terceira coluna.

Assim, o índice α determina quanto a isofota de Petrosian se afasta de uma elipse; daí o seu nome.

Imaginamos que a isofota de Petrosian seja uma corda. Quando se estica a corda para obter uma elipse, conservando a mesma elipticidade que a isofota de Petrosian, a área no interior aumenta. α mede esse aumento em termo relativo.

Como mencionado acima, α foi concebido para medir traços de perturbações gravitacionais. Não é um índice morfológico adaptado para diagnosticar os tipos morfológicos habituais. α não tem qualquer capacidade para separar os tipos e não mostra quase nenhuma evolução ao longo da sequência de Hubble (Figura 3.19). A série ligeiramente crescente entre Sa e Sd é real e corresponde a um aumento das regiões de formação estelar na periferia da galáxia que podem formar pequenas excrescências na isofota de Petrosian. A fase decrescente entre E e Sa é puramente técnica: na parte externa, o perfil de de Vaucouleurs é muito estendido e plano, o que cria uma isofota denteada por causa do ruído.

Em média, α é independente do tipo, porque as galáxias de todos os tipos morfológicos, das elípticas às “magalhânicas”, têm contornos regulares, desde que em equilíbrio gravitacional. Somente uma força externa à galáxia pode deformar a sua silhueta.

Comparando α com os índices precedentes (Figura 3.20), constatamos, sem surpresa, que as correlações são fracas ou inexistentes. Há, todavia, pequenas relações com os índices que quantificam irregularidades: α segue muito grosseiramente os comportamentos de A e S . Em particular, os cantos superior esquerdo e inferior direito do plano α - A são notavelmente desertos. É um argumento que consolida a nossa escolha de combinar esses dois índices para definir um indicador de galáxia em interação (Capítulo 5, Seção 5).

3. MORFOLOGIA QUANTITATIVA

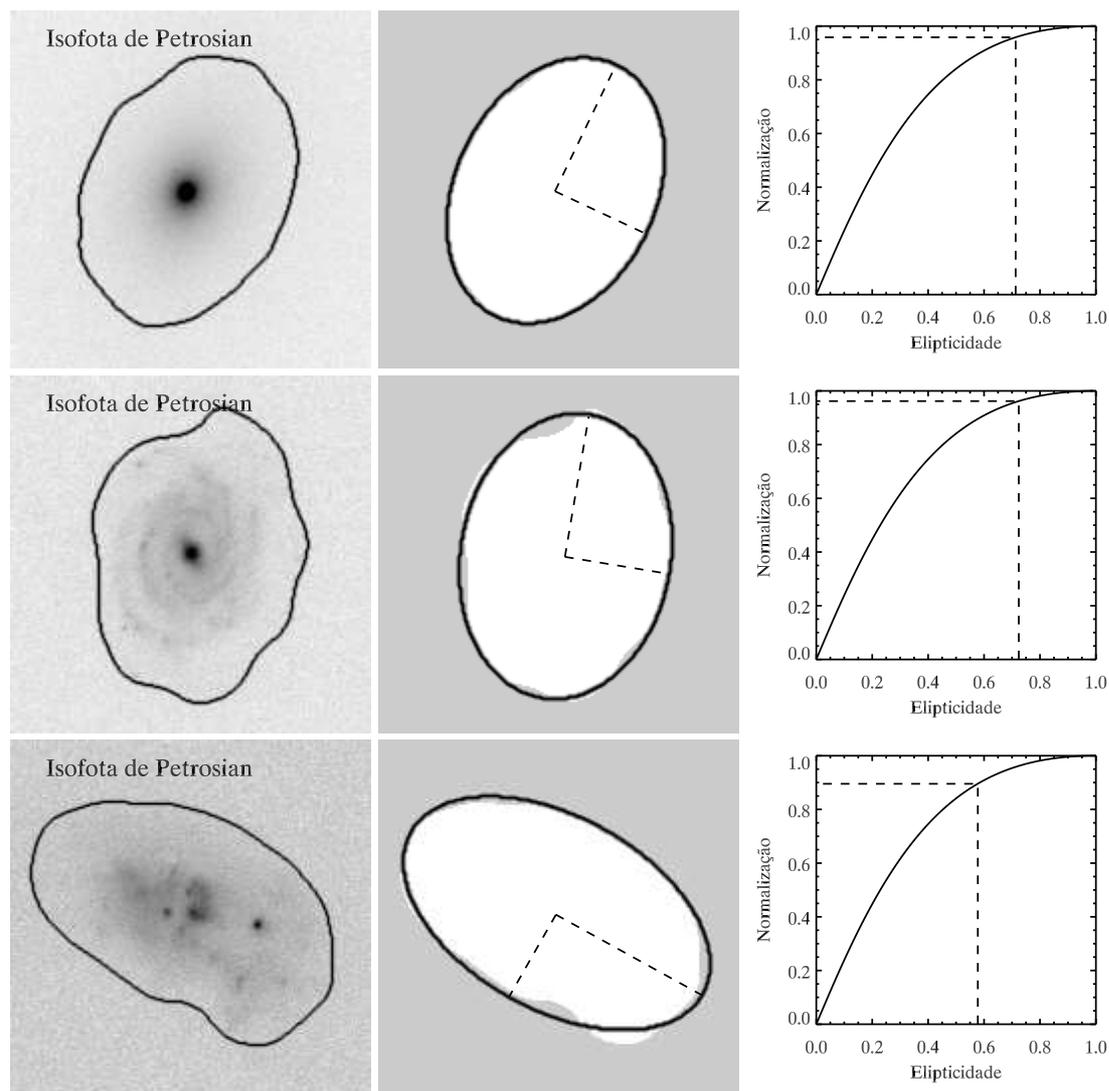


Figura 3.18: Processo de computação do índice de anelpticidade α .

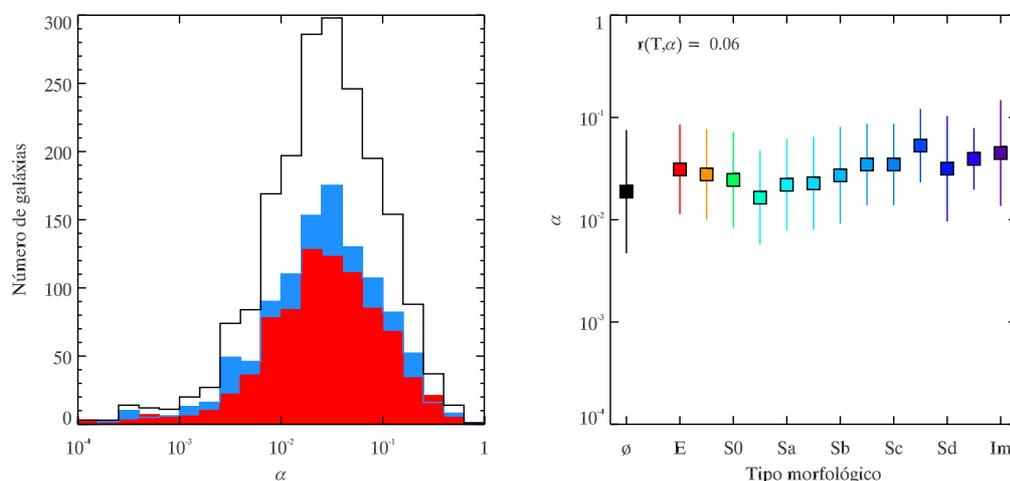


Figura 3.19: Distribuição de α (esquerda) e mediana de α em função do tipo morfológico (direita).

3.3.2 O índice de granulação

O índice de granulação visa medir como as regiões de formação estelar são distribuídas na galáxia. Comparamos a dispersão radial das estruturas de pequeno tamanho com aquela das estruturas de grande tamanho. Usamos, para este fim, uma ferramenta de análise multi-escala: a transformada de *wavelet*.

A palavra inglesa *wavelet* pode ser traduzida em português por ondinha. A transformada de *wavelet* é uma operação que decompõe um sinal em uma série de coeficientes nos quais a informação espacial é conservada. As funções de base, sobre as quais o sinal é projetado, são as próprias *wavelets*. Há várias famílias de *wavelet* possíveis, mas todas satisfazem as seguintes condições: são funções quadrado-somável (frequentemente definidas sobre um suporte finito), de média nula, e que apresentam uma oscilação, permitindo de definir uma pseudofrequência.

A *wavelet* não é uma função periódica, porque a oscilação não acontece em todo seu domínio de definição, mas apenas em um lugar. Isso implica a conservação da localização nos coeficientes da transformada de *wavelet*, ao contrário da transformada de Fourier. Da transformada de Fourier resultam duas funções – em geral designadas como uma única função complexa – no espaço das frequências: uma amplitude⁵ e

⁵A amplitude em função da frequência é chamada de espectro de frequência; o quadrado da ampli-

3. MORFOLOGIA QUANTITATIVA

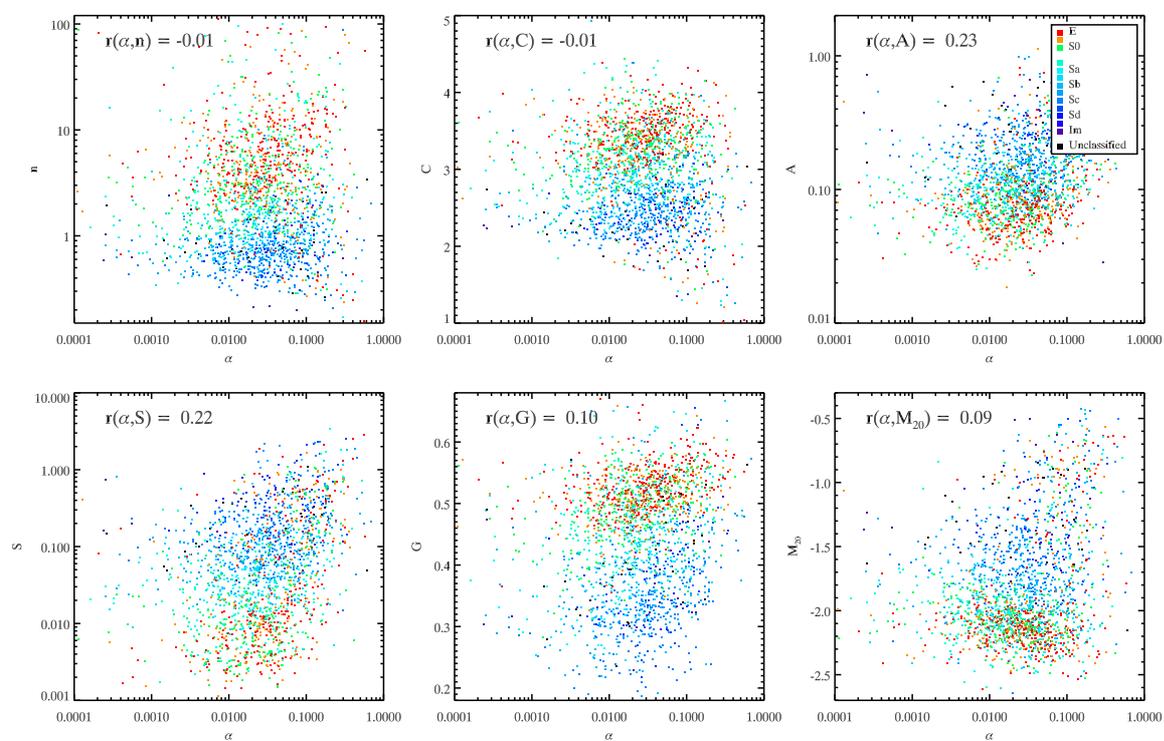


Figura 3.20: Relações entre α e os índices precedentes.

uma fase. Durante a transformada de *wavelet*, a localização da *wavelet* é uma variável de saída, como se as amplitudes fossem computadas para todas as fases possíveis. O resultado da transformada de *wavelet* é assim uma só função de duas variáveis: a frequência e a localização, grandeza de mesma unidade do que o sinal de entrada. A transformada de *wavelet* é altamente redundante, pois o sinal de saída tem uma dimensão adicional em relação ao sinal de entrada. É por isso que a transformada de *wavelet* discreta é computada apenas para algumas frequências distintas, as potências de 2.

Adaptando a metodologia para um sinal espacial bidimensional, ou seja, uma imagem, a *wavelet* se torna uma função da posição e não do tempo. Nessas condições, falar de frequência é um abuso de linguagem; encontra-se, às vezes, a expressão frequência espacial. Em nosso trabalho, computamos a transformada de *wavelet* usando o algoritmo *à trous* apresentado no Apêndice do Capítulo 5. Os coeficientes de *wavelet* são imagens do mesmo tamanho que a imagem original. Uma *wavelet* atua como um filtro passa-faixa, de modo que os coeficientes contêm apenas os elementos do tamanho da escala da *wavelet*.

Resumindo, a transformada de *wavelet* separa a imagem original em uma família de imagens de acordo com o tamanho das estruturas. Somando todos os coeficientes, recupera-se a imagem original.

A primeira coluna da Figura 3.21 mostra a transformada de *wavelet* da galáxia espiral *Fukugita0443*. O primeiro coeficiente W_1 é composto das estruturas de 2 píxeis; o segundo coeficiente W_2 é composto das estruturas de 4 píxeis, W_3 de 8 píxeis, e assim por diante. Em cada coeficiente, as zonas de interesse são determinadas por um *threshold* a 3σ . Fora das zonas de interesse, os píxeis são ignorados.

A distribuição radial pode ser medida pelo momento centrado de segunda ordem. Multiplicamos os coeficientes de *wavelet* com o quadrado da distância ao centro galáctico: Wr^2 (Figura 3.21, coluna da direita). Para considerar a forma geral da galáxia e a sua elipticidade, r não é o raio de um círculo, mas o raio da elipse, expresso em fração do raio total.

O espectro dos momentos de segunda ordem é construído somando as intensidades em cada uma das imagens $W_n r^2$. Esse fluxo total é expresso como uma magnitude,

tude é chamado de espectro de potência ou densidade espectral de potência.

3. MORFOLOGIA QUANTITATIVA

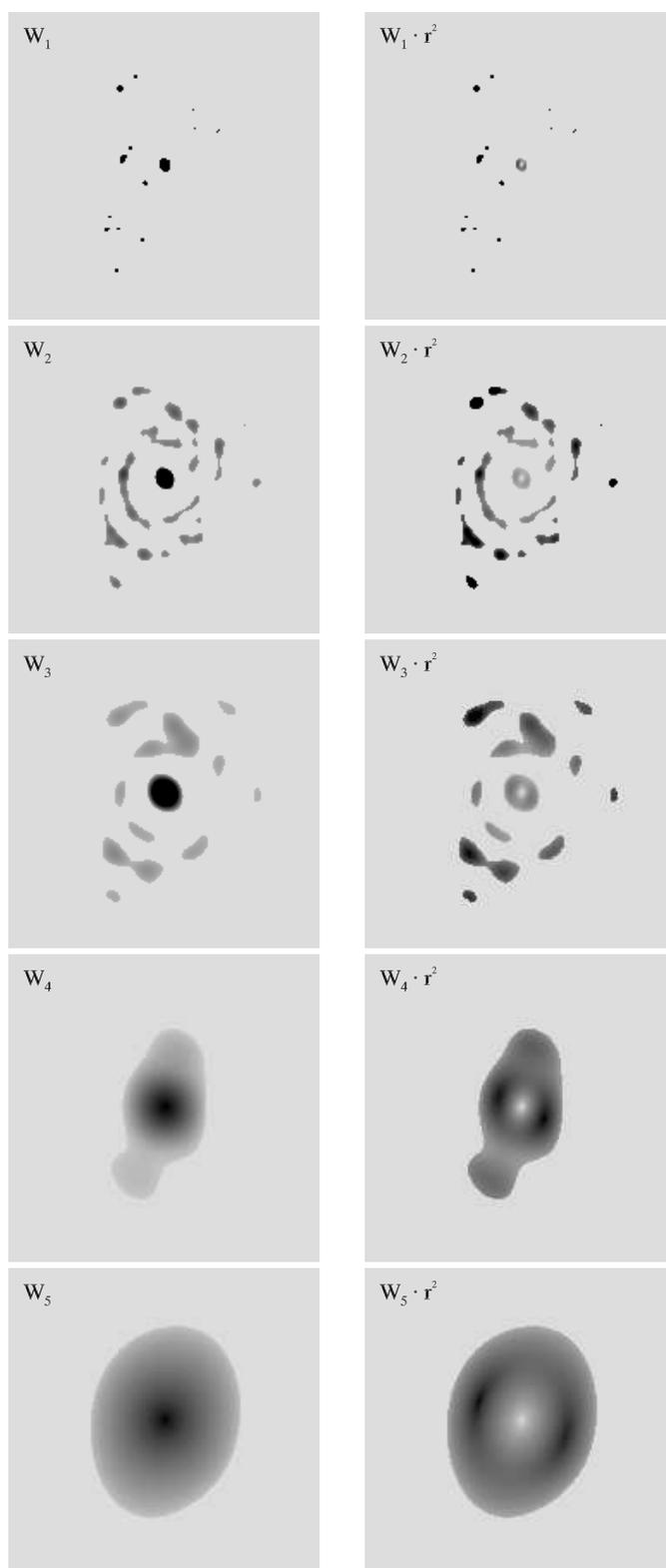


Figura 3.21: Esquerda: Decomposição em *wavelets* da galáxia *Fukugita0443*. Direita: Os coeficientes de *wavelet* são multiplicados por r^2 para obter o momento central de segunda ordem. Os fluxos totais em cada uma destas imagens formam o espectro de potência.

3.3 Dois novos índices morfológicos

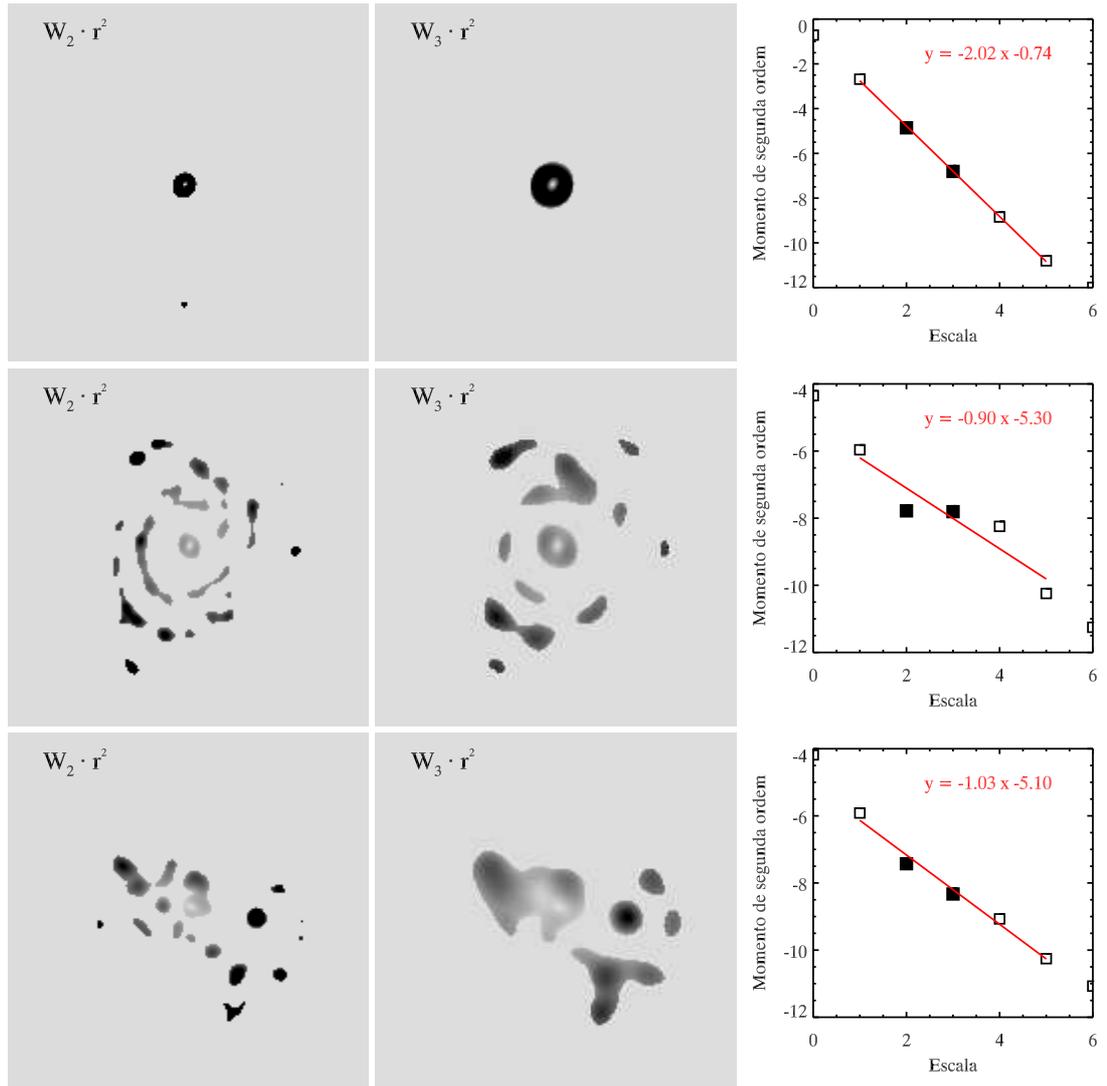


Figura 3.22: Processo de computação do índice de granulação γ .

3. MORFOLOGIA QUANTITATIVA

ou seja, em uma escala logarítmica inversa. Na Figura 3.22 são representados os coeficientes ponderados das escalas 2 (W_2r^2 , primeira coluna) e 3 (W_3r^2 , segunda coluna) para as nossas galáxias exemplos. Os fluxos totais dessas imagens, em unidade de magnitude, são indicados pelos quadrados negros nos gráficos da terceira coluna. Os pontos das outras escalas são desenhados como quadradinhos brancos, completando o espectro.

Os espectros de todas as galáxias são, em média, decrescentes. Isso demonstra – pois a escala do espectro é invertida – que as pequenas estruturas são mais centradas e menos luminosas que as grandes. As estruturas de grandes escalas, tipicamente formadas pelo disco, contribuem, em geral, para a maior parte do fluxo e se espalham até a borda da galáxia.

O índice de granulação γ é definido como a inclinação média do espectro, avaliada por uma regressão linear (reta vermelha nos gráficos da Figura 3.22). O espectro é uma medida cruzada da intensidade e da dispersão radial das estruturas, conforme a sua escala. A granulação γ ficará, então, pequena para galáxias cujas estruturas de pequenas escalas estão no centro. É o caso das elípticas, nas quais o centro muito agudo e intenso é revelado como uma estrutura de pequena escala pela decomposição em *wavelets*, mas atenuado pelo raio quase nulo na região central.

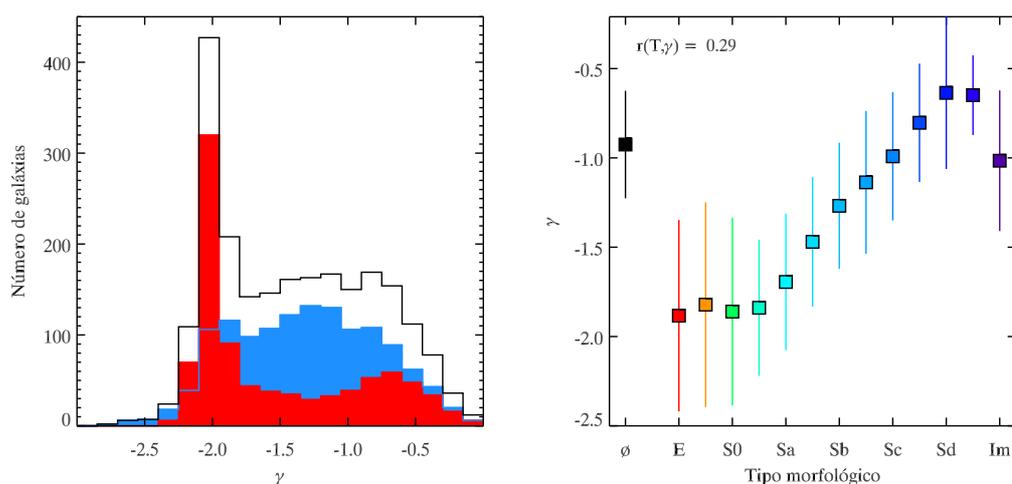


Figura 3.23: Distribuição de γ (esquerda) e mediana de γ em função do tipo morfológico (direita).

Na Figura 3.23, a distribuição de γ para as galáxias *early-type* exibe um pico muito marcado em $\gamma=-2$ (histograma vermelho). Porém, a transformada de *wavelet* é uma ferramenta tão poderosa para a análise multiescala que o mais fraco dos picos luminosos é detectado. Com o peso estatístico do quadrado do raio, um objeto, mesmo que pouco luminoso, adquire uma importância considerável quando se encontra na periferia. É o que acontece com as numerosas galáxias *early-type* que apresentam uma forte granulação ($\gamma > -1.5$).

No entanto, é importante notar que os *clumps* observados nas galáxias *early-type* não são sempre artefatos. É certamente possível que uma estrela do primeiro plano ou uma galáxia do fundo se encontre na isofota de Petrosian e atrapalhe a análise. Mas, outras vezes, trata-se de uma galáxia companheira ou satélite. Nestes casos, γ pode realmente alcançar valores elevados.

Do lado direito da Figura 3.23, a relação entre γ e o tipo morfológico mostra o mesmo regime duplo que o índice M_{20} (Seção 3.2.6). No início da sequência de Hubble, entre E e Sa, γ é quase constante e tem grandes barras de dispersão. Para aumentar a eficiência de uma classificação automática, γ deveria ser associado pelo menos a um índice do tipo concentração, muito eficiente nessa zona. Inversamente, a correspondência entre o tipo e γ apresenta uma regularidade surpreendente entre Sa e Sd.

Na Figura 3.24, γ é comparado aos sete outros índices morfológicos. De maneira geral, encontramos de novo a distribuição estreita da figura precedente. Nos planos $\gamma-n$, $\gamma-C$, ou $\gamma-G$, a nuvem de pontos enche de maneira eficiente o espaço e o gradiente de cores muito suave deixa prever uma separação dos tipos morfológicos melhor sucedida que com qualquer outra combinação. No plano $\gamma-G$, há pouca poluição de uma cor na outra, embora as galáxias elípticas (pontos vermelhos e laranja) e lenticulares (pontos verdes) ocupassem a mesma região. Sem dúvida, o par $\gamma-G$ deve entrar em uma classificação automática.

O comportamento estranho visível no plano $\gamma-S$ é estudado em detalhe na Seção 4.2 do Capítulo 5. Durante a computação de S , o resíduo é próximo de um coeficiente de *wavelet* de pequena escala e o fato de esconder o centro age como um peso estatístico radial. Embora muito mais básica, a metodologia do cálculo de S é um pouco semelhante àquela de γ . Isso explica por que esses dois índices medem parcialmente a mesma coisa para as galáxias *late-type*.

3. MORFOLOGIA QUANTITATIVA

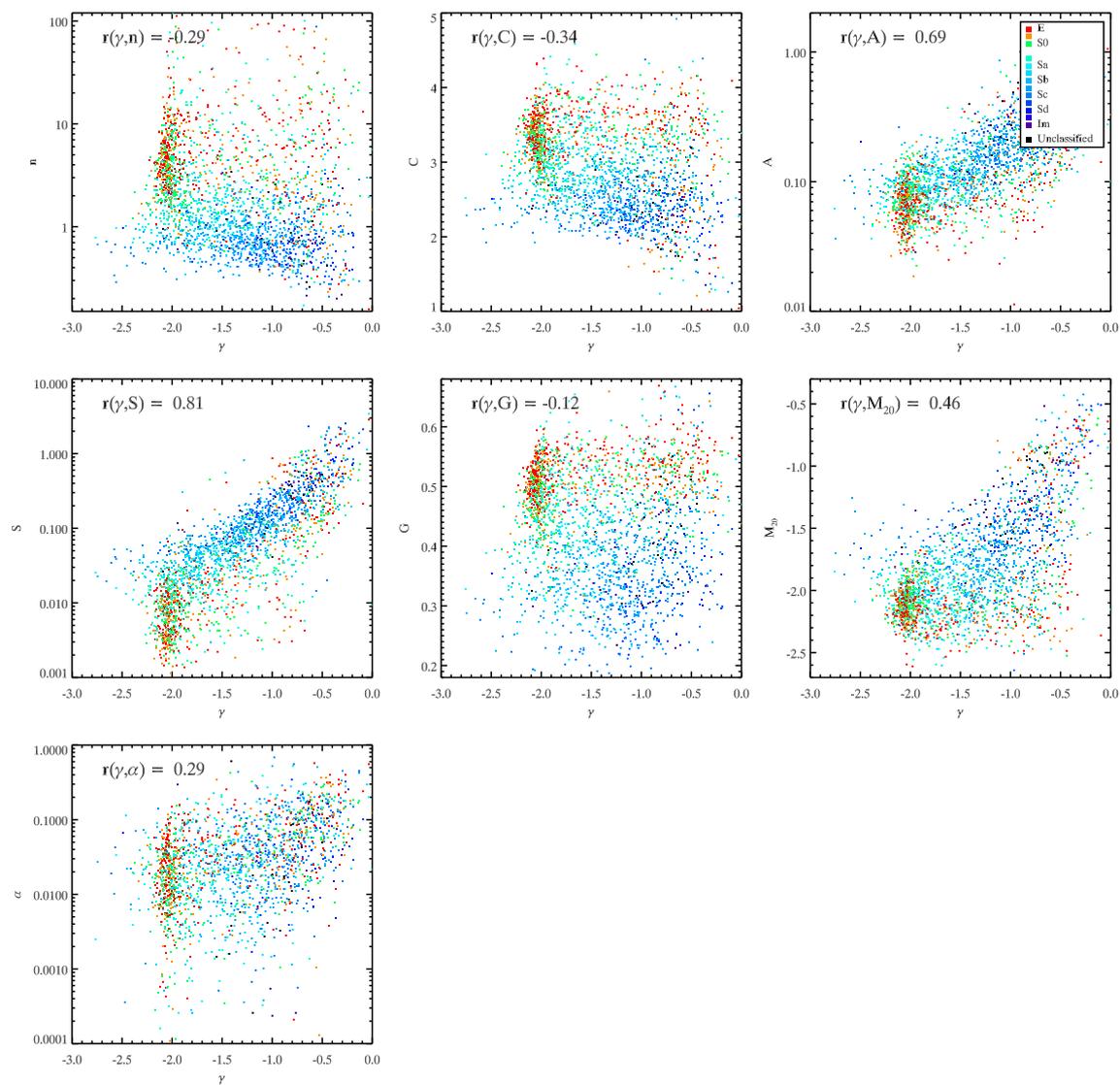


Figura 3.24: Relações entre γ e os índices precedentes.

Nesse capítulo, fizemos uma retrospectiva completa dos índices morfológicos comumente encontrados na literatura. Introduzimos também a anelipticidade e a granulação, dois novos índices concebidos particularmente para estudar as interações galácticas. Apresentamos todas as etapas dos processos de computação e os conceitos por trás das metodologias. Em seguida, analisamos a relação entre os índices e o tipo de Hubble, bem como as suas competências para separar as galáxias *early-type* e *late-type*. Finalmente, examinamos os índices dois a dois e as correlações entre si.

3. MORFOLOGIA QUANTITATIVA

Capítulo 4

Classificações morfológicas automáticas

4.1 Introdução

Chama-se de classificação automática o conjunto dos métodos estatísticos, informáticos e computacionais que consiste em atribuir uma classe para objetos homogêneos. É um caso particular de aprendizagem automática (*machine learning*, em inglês), ela mesma sendo um campo de estudo da inteligência artificial.

Existem duas categorias de aprendizagens automáticas:

1. A aprendizagem supervisionada, também chamada de aprendizagem baseada em instância (*instance-based learning*, em inglês), na qual as classes são predefinidas. A partir de um conjunto de instâncias, cujas classes já são conhecidas, um algoritmo prevê a afiliação dos objetos para classificar a uma das classes do modelo. Principais algoritmos: regressão linear, modelo de mistura gaussiana, máquinas de vetores de suporte (*support vector machines*), rede neural (*neural network*), *k* vizinhos mais próximos (*k-nearest neighbour*). No contexto estatístico, fala-se de análise discriminante.
2. A aprendizagem não-supervisada, na qual as classes são indefinidas. O sistema não dispõe de um conjunto de instâncias. O número, a natureza e as propriedades das classes são determinados, analisando diretamente os dados do sistema. No contexto estatístico, fala-se de *clustering* (agrupamento, aglomeração).

4. CLASSIFICAÇÕES MORFOLÓGICAS AUTOMÁTICAS

De maneira geral, uma classificação automática é caracterizada por três perguntas: Se tiverem, quais são as classes predeterminadas? Quais são as variáveis preditivas (variáveis de entrada)? Qual é o algoritmo escolhido?

Em Astronomia, as classificações morfológicas usam geralmente classes predeterminadas, dependendo do objectivo procurado: separação estrela-galáxia, separação entre *early-types* e *late-types*, galáxias barradas ou não, tipo de Hubble, outros tipos de galáxias em *redshift* mais alto, etc... As variáveis preditivas também podem ser de natureza diversa. O mais coerente é usar os índices morfológicos, pois foram desenvolvidos para coincidir com a morfologia. Mas, frequentemente, usa-se outros parâmetros galácticos relacionados com a morfologia, tais como a cor, o diâmetro, a magnitude em várias bandas, o brilho superficial em várias aberturas e bandas, a elongação, o gradiente do brilho superficial, o gradiente de cor ou a fração do fluxo no bojo.

Existe uma literatura abundante sobre as classificações morfológicas em função das numerosas técnicas possíveis. Por exemplo, Huertas-Company et al. (2008) procuram separar as galáxias *early-type* e *late-type* aplicando o algoritmo das máquinas de vetores de suporte sobre os índices morfológicos. Storrie-Lombardi et al. (1992) estimam o tipo de Hubble pelo algoritmo da rede neural sobre uma grande quantidade de medidas fotométricas. Scarlata et al. (2007) empregam a *PCA* sobre os índices morfológicos para determinar o tipo de Hubble. Ao separar as galáxias elípticas, espirais e *edge-on*, Shamir (2009) usa o algoritmo dos vizinhos mais próximos sobre um número muito grande de características de imagem computadas de maneiras muito diversas.

Em lugar de um algoritmo de aprendizagem supervisionada, a separação entre as classes é, às vezes, reduzida a um processo rudimentar, como quando é realizada visualmente no espaço dos parâmetros. A fim de identificar galáxias em fusão, Conselice (2003); Conselice et al. (2003) determinam visualmente um linha reta na classificação *CAS*; de mesma maneira que Lotz et al. (2004) no plano $G-M_{20}$. Ou ainda Park & Choi (2005), no plano cor-gradiente de cor, para separar as galáxias *early-type* das *late-type*.

Andrae et al. (2011a) passam em revista os métodos usados em recentes classificações morfológicas automáticas. Destacam potenciais e armadilhas dos diferentes parâmetros preditivos e de várias técnicas. Pela nossa parte, as variáveis preditivas são os índices morfológicos do Capítulo 3, e as classes predefinidas são os tipos de Hubble. Procuramos avaliar o papel desempenhado pela escolha do algoritmo sobre a qualidade da classificação.

4.2 As nossas classificações

O objetivo das nossas classificações automáticas é reproduzir a sequência de Hubble, tal como foi quantificada no artigo de Fukugita et al. (2007). Estamos, então, na situação de aprendizagem baseada em instância. Em tais métodos, uma amostra de instâncias de treinamento já foi anteriormente classificada por uma técnica diferente; aqui, uma classificação visual. Em seguida, um algoritmo compara os objetos estudados com as instâncias de treinamento.

A amostra de treinamento consiste de vetores de entrada e vetores de resposta. No nosso caso, o vetor de entrada é formado dos índices morfológicos e o vetor de resposta do tipo de Hubble. A classificação das instâncias de treinamento é considerada como exata. Essa suposição é, evidentemente, apenas uma hipótese de trabalho. Em nosso catálogo morfológico, o tipo foi estimado por cada um dos três primeiros autores do artigo. O valor retido T é a média arredondada, mas é indicado também, como medida da incerteza, o desvio padrão dos três valores. Em média sobre o catálogo todo, o desvio padrão vale $\langle\sigma_T\rangle=0.37$. Isso ilustra que, mesmo dentro de um grupo se concertando, uma classificação é afetada pela subjetividade dos classificadores.

De uma pesquisa para a outra, as diferenças ainda são maiores. Usando também imagens do SDSS, Nair & Abraham (2010) classificaram visualmente 14 000 galáxias. Eles comparam a sua classificação com aquelas similares do *Third Reference Catalogue of Bright Galaxies – RC3* (de Vaucouleurs et al. 1991) e de Fukugita que usamos em nosso trabalho. Os gráficos, tirados do artigo original, são reproduzidos na Figura 4.1. Nesse gênero de representação, às vezes chamada de histograma bidimensional, o tamanho dos círculos é proporcional ao número de objetos localizados em um lugar. Os desvios médios entre as classificações são de $\langle|T_{PN} - T_{RC3}| \rangle \approx 1.2$ para RC3 (nota-se que a escala usada para medir T é diferente, quase dobrada) e $\langle|T_{PN} - T_{Fukugita}| \rangle \approx 0.8$ para Fukugita.

Todas as classificações contêm um erro intrínseco, devido ao espalhamento real das propriedades observadas. Em uma classificação automática, se adicionam também as incertezas de medidas sobre os índices e as imprecisões do modelo usado pelo algoritmo. Podemos, portanto, aceitar o desvio médio de 0.8 como valor de referência para avaliar as nossas classificações.

4. CLASSIFICAÇÕES MORFOLÓGICAS AUTOMÁTICAS

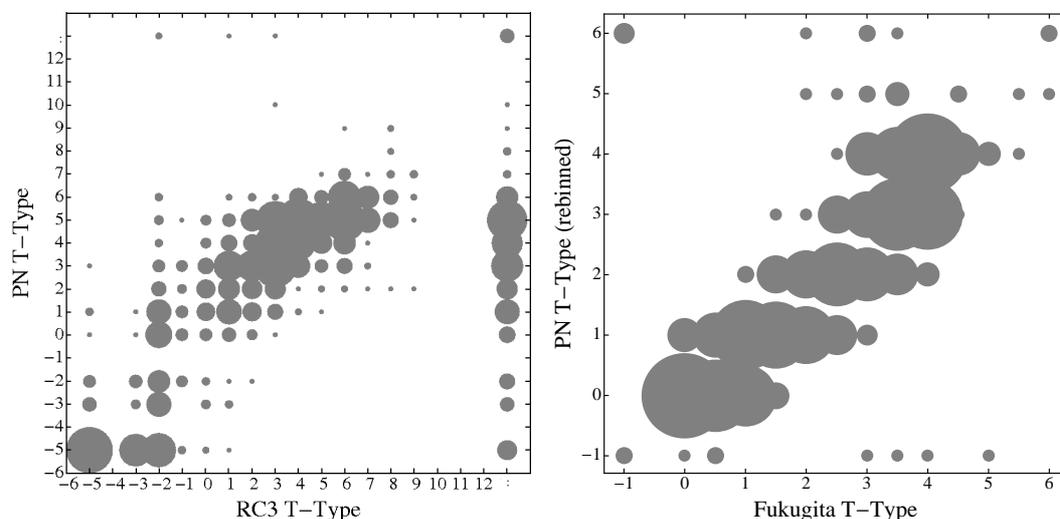


Figura 4.1: Comparação da classificação visual efetuada por Nair (notada de PN) com aquelas de RC3 (esquerda) e de Fukugita (direita). Figura tirada de Nair & Abraham (2010).

Uma classificação automática aproveita a informação contida no vetor de entrada para determinar o vetor de resposta. Pode-se pensar que quanto maior o número de parâmetros de entrada, melhor sucedida se torna a classificação. Todavia, o número de configurações de um sistema é exponencialmente proporcional à sua dimensão, de tal modo que aumentar cegamente o número de parâmetros no vetor de entrada pode alterar sensivelmente a capacidade de computação.

No Capítulo 3, mostramos que certos índices têm comportamentos similares e que outros são complementários. Uma primeira família de índices são de tipo concentração: n , C e G . Dos três, o mais abrangente, sem obrigação de definir um centro ou de assumir uma simetria de rotação, e igualmente o mais rápido para computar, é G . Conservamo-lo. Uma segunda família de índices tenta avaliar a textura da galáxia: S , γ e A . Os dois primeiros são muito correlacionados, e preferimos γ porque as galáxias *early-type* são confinadas ali em um intervalo pequeno. De seu lado, A mede uma propriedade bem específica de que é difícil se dispensar. M_{20} é, de uma certa maneira, um índice híbrido e revela uma concordância notável com o tipo morfológico; conservamo-lo igualmente. Por sua vez, α não demonstra nenhuma aptidão para descrever a sequencia de Hubble.

Por conseguinte, temos como vetor de entrada do nosso sistema de classificação automática:

$$x = (G, \gamma, \log(A), M_{20}). \quad (4.1)$$

Neste Capítulo, descartamos do catálogo morfológico as galáxias não classificadas na sequência de Hubble ($T=-1$). Além disso, aplicamos o critério de qualidade fotométrica discutido no Capítulo 5, onde o estudo da robustez dos índices morfológicos mostrou que as galáxias insuficientemente resolvidas fornecem medidas inconfiáveis. Então, as galáxias, cujo diâmetro angular é menor que oito vezes o *seeing*, são também descartadas. Das 2253 galáxias no catálogo de Fukugita, 1793 (79.6%) entram agora em nossa amostra de trabalho.

Independentemente do método escolhido, a amostra de treinamento deve ser composta de instâncias para as quais os resultados são indubitáveis. Construímos, para este fim, uma amostra apurada na qual os pontos divergentes são excluídos. Os pontos divergentes não são aqui considerados como medidas erradas, mas como valores atípicos, excessivamente afastados da tendência central para ser representativos da maioria. Os pontos divergentes são diagnosticados seguindo o método exposto no Apêndice A: (1) em cada tipo, os componentes principais são determinados com as pseudo-covariâncias; (2) determina-se a hiperelipse, centrada nas medianas e cujos eixos são os PCs de comprimento $3\bar{\sigma}$; (3) os pontos fora da hiperelipse são excluídos; (4) o procedimento é iterado até que nenhum ponto seja descartado. A amostra apurada contém 1181 galáxias (65.8% da amostra de trabalho) e é usada como amostra de treinamento.

No espaço quadridimensional da variável x , desenvolvemos quatro métodos de classificação: a aproximação linear, a hipótese de distribuição normal multivariada, a determinação empírica de uma distribuição qualquer, e as k -médias. Além de uma amostra de treinamento, o método requer uma amostra de controle para testar a validade da classificação. Em geral, não há nenhum problema em usar duas vezes o mesmo catálogo. Mas para o terceiro método, temos que separar a nossa amostra de trabalho em duas sub-amostras, uma de treinamento e uma de controle.

4. CLASSIFICAÇÕES MORFOLÓGICAS AUTOMÁTICAS

4.2.1 Regressão linear

A primeira classificação é um subterfúgio simples para transformar o problema de análise discriminante em um problema de regressão linear múltipla. Efetivamente, em nosso caso, a variável a prever – o tipo morfológico T – não é realmente uma variável discreta. Usamos o fato que a sequência de Hubble seja ordenada e quantificada para exprimir T como uma combinação linear dos índices morfológicos.

Nesse método, a sequência de Hubble é suposta exibir um comportamento linear de acordo com os índices. No espaço de cinco dimensões formadas das quatro variáveis de entrada adicionadas do tipo morfológico, as galáxias se espalham ao redor de um hiperplano. A equação do hiperplano é determinada pelo método dos mínimos quadrados. Comparando o tipo assim modelizado com o tipo de Fukugita (Figura 4.2, esquerda), notamos que o ajuste favorece os tipos morfológicos encontrados no centro da sequência de Hubble e que aqueles nas extremidades são sub-representados. Particularmente, nenhuma galáxia é classificada como elíptica ($T=0$) ou irregular tipo Magalhães ($T=6$). A regressão linear direta, ajustando $y(x)$, tem tendência a restituir uma sequência compacta e centrada demais, o que se visualiza pela nuvem de pontos não oblíquo o suficiente de acordo com a diagonal.

A fim de recobrar os tipos extremos, queríamos girar a nuvem de pontos até sobrepor a diagonal. Para isso, aplicamos uma segunda regressão, essa vez a uma dimensão e inversa, ou seja, ajustando $x(y)$. A segunda regressão aumenta o desvio médio entre o tipo ajustado e o tipo real, tornando a classificação menos eficiente conforme esse critério. Apesar disso, preferimos o método com a dupla regressão porque aparece menos enviesado (Figura 4.2, esquerda).

O estimador linear T_{linear} do tipo de Hubble pode ser expresso como uma combinação linear dos quatro índices morfológicos escolhidos. Com a dupla regressão, obtemos:

$$T_{linear} = -7.88 \cdot G + 1.42 \cdot \gamma + 1.37 \cdot \log(A) + 0.67 \cdot M_{20} + 10.46 \quad (4.2)$$

A Tabela 4.1 indica os resultados da classificação por regressão linear T_{linear} comparados à classificação de Fukugita $T_{Fukugita}$ para as 20 primeiras galáxias da amostra de trabalho. Como o problema inicial de análise discriminante foi transformado em um ajustamento, T_{linear} pode tomar qualquer valor, não apenas números semi-inteiros como é o caso na classificação de Fukugita.

4.2 As nossas classificações

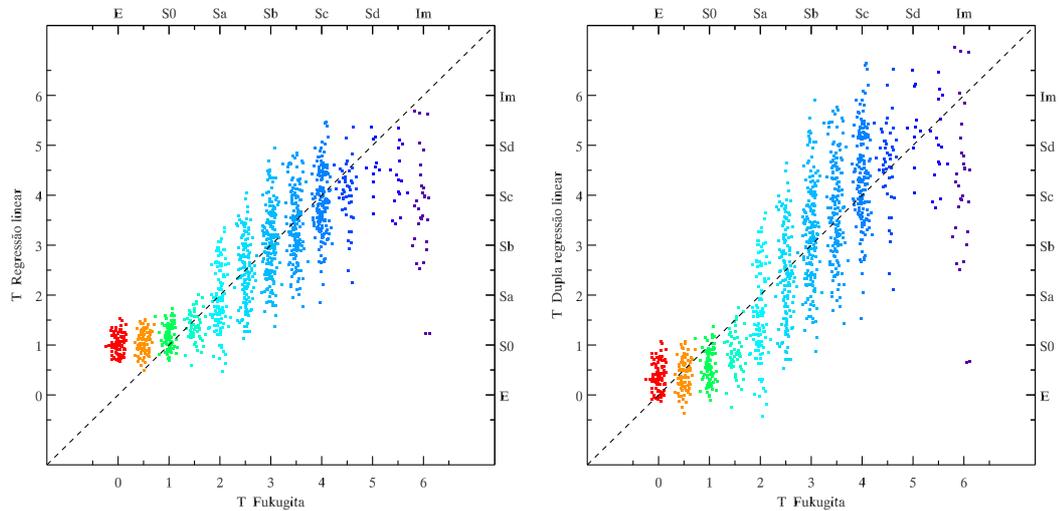


Figura 4.2: Esquerda: Comparação entre o tipo morfológico determinado por uma regressão linear e o tipo de referência. Direita: Uma segunda regressão, inversa, ajusta novamente o tipo morfológico. A posição dos pontos em X é ligeiramente deslocada para mais clareza. A linha na diagonal indica a igualdade.

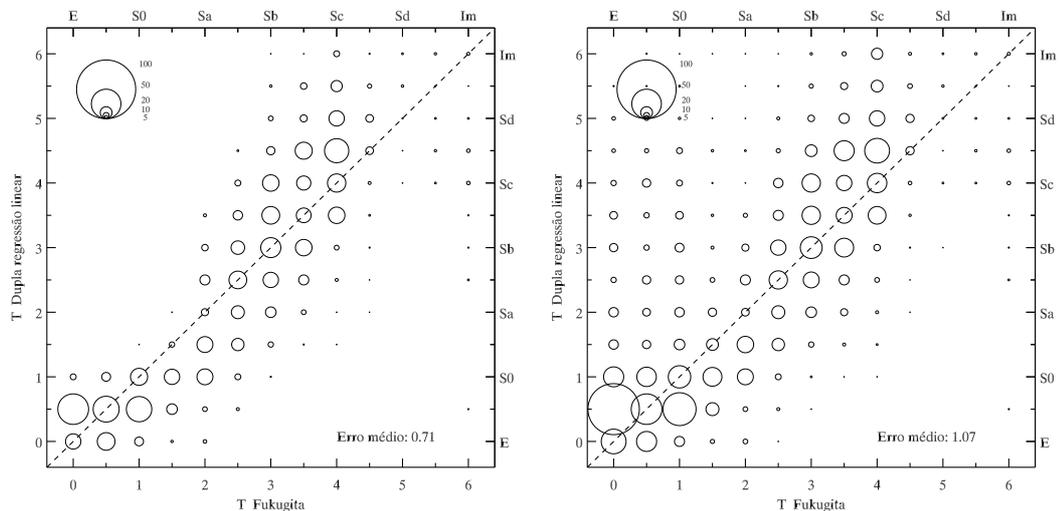


Figura 4.3: Comparação do tipo determinado por regressão e a classificação de Fukugita para a amostra apurada (esquerda) e a amostra de trabalho (direita). O tamanho dos círculos é proporcional ao número de galáxias.

4. CLASSIFICAÇÕES MORFOLÓGICAS AUTOMÁTICAS

Tabela 4.1: Tipos morfológicos pelo método da dupla regressão linear.

ID	$T_{Fukugita}$	T_{linear}
Fukugita0001	0.5	1.07
Fukugita0002	3.0	2.40
Fukugita0003	1.0	0.84
Fukugita0004	3.5	4.79
Fukugita0005	3.5	3.09
Fukugita0006	0.0	0.05
Fukugita0007	4.0	5.19
Fukugita0009	3.0	2.13
Fukugita0010	2.0	0.32
Fukugita0012	3.5	2.80
Fukugita0013	4.0	3.86
Fukugita0016	1.0	3.11
Fukugita0017	4.0	2.14
Fukugita0019	1.0	1.83
Fukugita0021	0.5	1.47
Fukugita0022	6.0	4.85
Fukugita0023	1.0	1.22
Fukugita0024	1.5	0.67
Fukugita0025	2.5	2.81
Fukugita0026	6.0	3.02

Na Figura 4.3, comparamos a classificação linear e aquela de Fukugita para a amostra apurada (esquerda) e a amostra de trabalho (direita). O tamanho dos círculos é proporcional ao número de galáxias, tal como indicado pela escada no recanto superior esquerdo do gráfico. Vemos claramente o efeito da filtragem que engendrou a amostra apurada. As galáxias que foram descartadas são, de maneira geral, muito menos bem classificadas que aquelas da amostra apurada. Temos um erro médio $\langle |T_{linear} - T_{Fukugita}| \rangle = 0.71$ para a amostra apurada e 1.07 para a amostra de trabalho. Comparados com $\langle |\Delta T| \rangle = 0.8$ das duas classificações visuais citadas em introdução, esses valores ilustram que um algoritmo simples alcança resultados parecidos àqueles das classificações visuais, para galáxias convenientes. Mas a classificação automática não consegue atribuir um tipo certo, desde que a imagem apresente uma pequena anomalia fotométrica.

4.2.2 Distribuição normal multivariada

A nossa segunda classificação automática é um método probabilístico, quer dizer, que faz-se uma hipótese sobre a distribuição dos objetos para classificar.

Uma hipótese razoável é dizer que, em cada tipo morfológico, os índices morfológicos seguem uma distribuição normal. Nesse caso, fala-se de modelo de mistura gaussiana. É o processo usado em duas dimensões no Apêndice A, quando a região ocupada por um tipo é estimada por uma elipse. A distribuição normal multivariada é um modelo simples de *clustering*. No espaço quadridimensional dos parâmetros, as galáxias de um mesmo tipo se reúnem em um grupo compacto e regular, cuja densidade $f(x)$ pode ser estimada por uma função gaussiana multidimensional:

$$f(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(x - \langle x \rangle)^T \Sigma^{-1} (x - \langle x \rangle)\right), \quad (4.3)$$

onde n é a dimensão da variável, ou seja, aqui $n=4$; $\langle x \rangle$ é a média da variável x ; e Σ é a matriz de covariância $\Sigma = \text{Cov}(x)$, $|\Sigma|$ é seu determinante e Σ^{-1} sua matriz inversa.

A densidade $f(x)$ medida sobre a amostra das instâncias é tratada como a uma densidade de probabilidade de encontrar um certo tipo em um certo lugar do espaço

4. CLASSIFICAÇÕES MORFOLÓGICAS AUTOMÁTICAS

dos parâmetros. A Equação 4.3 é inteiramente determinada pelas média e matriz de covariância. É desnecessário tomar precauções particulares, usando medianas e pseudo-covariâncias, porque a amostra de treinamento já é livre de eventuais pontos divergentes. Sobre a amostra apurada, então, calculamos a média $\langle x \rangle$ e a matriz de covariância Σ para cada tipo morfológico. Obtemos:

$$\langle x \rangle_{T=0.0} = \begin{pmatrix} 0.5180 \\ -2.0602 \\ -1.1839 \\ -2.1660 \end{pmatrix} \quad \Sigma_{T=0.0} = \begin{pmatrix} 0.0007 & 0.0006 & 0.0004 & -0.0016 \\ 0.0006 & 0.0020 & 0.0006 & -0.0006 \\ 0.0004 & 0.0006 & 0.0149 & 0.0020 \\ -0.0016 & -0.0006 & 0.0020 & 0.0079 \end{pmatrix} \quad (4.4)$$

$$\langle x \rangle_{T=0.5} = \begin{pmatrix} 0.5157 \\ -2.0489 \\ -1.1738 \\ -2.1724 \end{pmatrix} \quad \Sigma_{T=0.5} = \begin{pmatrix} 0.0008 & 0.0008 & 0.0001 & -0.0015 \\ 0.0008 & 0.0040 & 0.0006 & -0.0009 \\ 0.0001 & 0.0006 & 0.0210 & 0.0026 \\ -0.0015 & -0.0009 & 0.0026 & 0.0073 \end{pmatrix} \quad (4.5)$$

$$\langle x \rangle_{T=1.0} = \begin{pmatrix} 0.5056 \\ -2.0481 \\ -1.1203 \\ -2.1540 \end{pmatrix} \quad \Sigma_{T=1.0} = \begin{pmatrix} 0.0011 & 0.0013 & 0.0012 & -0.0018 \\ 0.0013 & 0.0070 & 0.0033 & -0.0002 \\ 0.0012 & 0.0033 & 0.0258 & 0.0003 \\ -0.0018 & -0.0002 & 0.0003 & 0.0082 \end{pmatrix} \quad (4.6)$$

$$\langle x \rangle_{T=1.5} = \begin{pmatrix} 0.4899 \\ -1.9829 \\ -1.0948 \\ -2.1626 \end{pmatrix} \quad \Sigma_{T=1.5} = \begin{pmatrix} 0.0019 & 0.0009 & 0.0011 & -0.0028 \\ 0.0009 & 0.0116 & 0.0001 & -0.0016 \\ 0.0011 & 0.0001 & 0.0214 & -0.0022 \\ -0.0028 & -0.0016 & -0.0022 & 0.0102 \end{pmatrix} \quad (4.7)$$

$$\langle x \rangle_{T=2.0} = \begin{pmatrix} 0.4561 \\ -1.6999 \\ -1.0189 \\ -2.0878 \end{pmatrix} \quad \Sigma_{T=2.0} = \begin{pmatrix} 0.0057 & 0.0096 & 0.0024 & -0.0077 \\ 0.0096 & 0.1738 & 0.0430 & 0.0051 \\ 0.0024 & 0.0430 & 0.0360 & 0.0114 \\ -0.0077 & 0.0051 & 0.0114 & 0.0292 \end{pmatrix} \quad (4.8)$$

4.2 As nossas classificações

$$\langle x \rangle_{T=2.5} = \begin{pmatrix} 0.4093 \\ -1.5274 \\ -0.9394 \\ -2.0249 \end{pmatrix} \quad \Sigma_{T=2.5} = \begin{pmatrix} 0.0055 & 0.0033 & 0.0009 & -0.0102 \\ 0.0033 & 0.1464 & 0.0319 & 0.0057 \\ 0.0009 & 0.0319 & 0.0241 & 0.0058 \\ -0.0102 & 0.0057 & 0.0058 & 0.0451 \end{pmatrix} \quad (4.9)$$

$$\langle x \rangle_{T=3.0} = \begin{pmatrix} 0.3674 \\ -1.3524 \\ -0.8314 \\ -1.8757 \end{pmatrix} \quad \Sigma_{T=3.0} = \begin{pmatrix} 0.0049 & 0.0031 & -0.0003 & -0.0125 \\ 0.0031 & 0.1511 & 0.0419 & 0.0202 \\ -0.0003 & 0.0419 & 0.0311 & 0.0133 \\ -0.0125 & 0.0202 & 0.0133 & 0.0699 \end{pmatrix} \quad (4.10)$$

$$\langle x \rangle_{T=3.5} = \begin{pmatrix} 0.3528 \\ -1.2077 \\ -0.7844 \\ -1.7332 \end{pmatrix} \quad \Sigma_{T=3.5} = \begin{pmatrix} 0.0045 & 0.0055 & 0.0006 & -0.0083 \\ 0.0055 & 0.1596 & 0.0489 & 0.0481 \\ 0.0006 & 0.0489 & 0.0351 & 0.0228 \\ -0.0083 & 0.0481 & 0.0228 & 0.0928 \end{pmatrix} \quad (4.11)$$

$$\langle x \rangle_{T=4.0} = \begin{pmatrix} 0.3242 \\ -1.0255 \\ -0.7422 \\ -1.5634 \end{pmatrix} \quad \Sigma_{T=4.0} = \begin{pmatrix} 0.0039 & 0.0040 & 0.0020 & -0.0073 \\ 0.0040 & 0.1286 & 0.0434 & 0.0451 \\ 0.0020 & 0.0434 & 0.0323 & 0.0140 \\ -0.0073 & 0.0451 & 0.0140 & 0.0772 \end{pmatrix} \quad (4.12)$$

$$\langle x \rangle_{T=4.5} = \begin{pmatrix} 0.3331 \\ -0.9241 \\ -0.6773 \\ -1.4904 \end{pmatrix} \quad \Sigma_{T=4.5} = \begin{pmatrix} 0.0044 & 0.0059 & 0.0021 & -0.0032 \\ 0.0059 & 0.1448 & 0.0375 & 0.0366 \\ 0.0021 & 0.0375 & 0.0294 & 0.0106 \\ -0.0032 & 0.0366 & 0.0106 & 0.0711 \end{pmatrix} \quad (4.13)$$

$$\langle x \rangle_{T=5.0} = \begin{pmatrix} 0.3340 \\ -0.6243 \\ -0.5419 \\ -1.2328 \end{pmatrix} \quad \Sigma_{T=5.0} = \begin{pmatrix} 0.0014 & 0.0022 & -0.0005 & -0.0030 \\ 0.0022 & 0.1015 & 0.0101 & 0.0650 \\ -0.0005 & 0.0101 & 0.0130 & 0.0137 \\ -0.0030 & 0.0650 & 0.0137 & 0.0831 \end{pmatrix} \quad (4.14)$$

4. CLASSIFICAÇÕES MORFOLÓGICAS AUTOMÁTICAS

$$\langle x \rangle_{T=5.5} = \begin{pmatrix} 0.3460 \\ -0.7625 \\ -0.5559 \\ -1.2990 \end{pmatrix} \quad \Sigma_{T=5.5} = \begin{pmatrix} 0.0072 & 0.0062 & 0.0041 & -0.0016 \\ 0.0062 & 0.0865 & 0.0231 & 0.0452 \\ 0.0041 & 0.0231 & 0.0326 & 0.0079 \\ -0.0016 & 0.0452 & 0.0079 & 0.0707 \end{pmatrix} \quad (4.15)$$

$$\langle x \rangle_{T=6.0} = \begin{pmatrix} 0.3826 \\ -1.0503 \\ -0.6710 \\ -1.2836 \end{pmatrix} \quad \Sigma_{T=6.0} = \begin{pmatrix} 0.0047 & -0.0059 & -0.0044 & -0.0109 \\ -0.0059 & 0.2397 & 0.1435 & 0.1368 \\ -0.0044 & 0.1435 & 0.1310 & 0.0823 \\ -0.0109 & 0.1368 & 0.0823 & 0.1391 \end{pmatrix}. \quad (4.16)$$

Em um problema de *clustering* idealmente caracterizado, as classes são claramente separadas umas das outras, deixando regiões vazias no espaço dos parâmetros. Mas não é o caso aqui, porque a sequência de Hubble é contínua e suave. As densidades de probabilidade se sobrepõem, às vezes, em grande parte.

Para prever o tipo de uma galáxia descrita pela variável x , calcula-se a densidade de cada tipo em x com a Equação 4.3. A probabilidade da galáxia de pertencer a um certo tipo é simplesmente a densidade $f(x)$ normalizada pelo total das densidades. O tipo atribuído à galáxia é aquele com a probabilidade máxima.

Na tabela 4.2, o tipo T_{normal} calculado com a distribuição normal multivariada fica na terceira coluna e pode ser comparado com o tipo de Fukugita $T_{Fukugita}$ da segunda coluna. As colunas seguintes indicam as probabilidades de cada tipo, os valores máximos sendo em negrito.

Tabela 4.2: Tipos morfológicos pelo método da distribuição normal multivariada.

ID	$T_{Fukugita}$	T_{normal}	$P_{T=0.0}$	$P_{T=0.5}$	$P_{T=1.0}$	$P_{T=1.5}$	$P_{T=2.0}$	$P_{T=2.5}$	$P_{T=3.0}$	$P_{T=3.5}$	$P_{T=4.0}$	$P_{T=4.5}$	$P_{T=5.0}$	$P_{T=5.5}$	$P_{T=6.0}$
Fukugita0001	0.5	2.0	0.00	0.00	0.00	0.27	0.56	0.14	0.01	0.00	0.00	0.00	0.00	0.00	0.01
Fukugita0002	3.0	2.5	0.00	0.00	0.00	0.00	0.11	0.39	0.30	0.14	0.04	0.01	0.00	0.00	0.01
Fukugita0003	1.0	1.0	0.03	0.31	0.44	0.21	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fukugita0004	3.5	5.0	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.08	0.11	0.13	0.47	0.15	0.03
Fukugita0005	3.5	2.5	0.00	0.00	0.00	0.00	0.08	0.26	0.21	0.19	0.20	0.04	0.00	0.00	0.03
Fukugita0006	0.0	0.5	0.23	0.57	0.16	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fukugita0007	4.0	4.5	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.13	0.22	0.38	0.18	0.03	0.00
Fukugita0009	3.0	6.0	0.00	0.00	0.00	0.00	0.17	0.19	0.06	0.12	0.02	0.02	0.00	0.01	0.41
Fukugita0010	2.0	1.5	0.00	0.08	0.15	0.64	0.09	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fukugita0012	3.5	3.0	0.00	0.00	0.00	0.00	0.02	0.26	0.49	0.17	0.01	0.04	0.00	0.00	0.02
Fukugita0013	4.0	3.0	0.00	0.00	0.00	0.00	0.02	0.11	0.25	0.21	0.23	0.12	0.05	0.03	0.00
Fukugita0016	1.0	2.0	0.00	0.00	0.00	0.00	0.29	0.13	0.16	0.19	0.08	0.07	0.00	0.05	0.03
Fukugita0017	4.0	2.0	0.00	0.00	0.00	0.00	0.57	0.25	0.16	0.02	0.00	0.00	0.00	0.00	0.00
Fukugita0019	1.0	2.0	0.00	0.00	0.00	0.00	0.57	0.26	0.07	0.05	0.01	0.00	0.00	0.00	0.05
Fukugita0021	0.5	2.0	0.00	0.00	0.00	0.00	0.97	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fukugita0022	6.0	5.0	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.08	0.10	0.14	0.43	0.17	0.06
Fukugita0023	1.0	1.0	0.08	0.14	0.39	0.25	0.08	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.01
Fukugita0024	1.5	1.5	0.00	0.01	0.04	0.82	0.11	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fukugita0025	2.5	5.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Fukugita0026	6.0	6.0	0.00	0.00	0.00	0.00	0.00	0.03	0.05	0.20	0.09	0.13	0.00	0.04	0.45

4. CLASSIFICAÇÕES MORFOLÓGICAS AUTOMÁTICAS

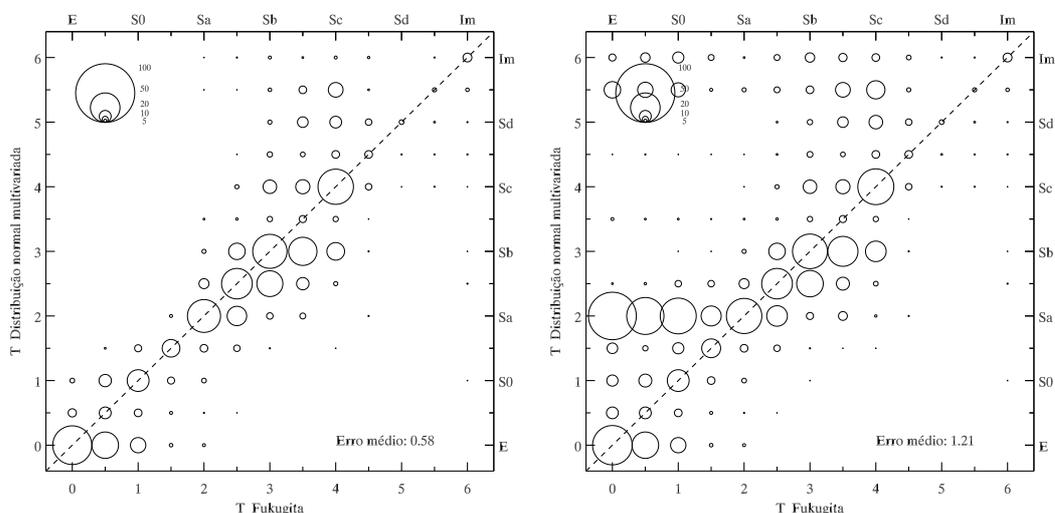


Figura 4.4: Comparação entre classificação pelo método da distribuição normal multivariada e o tipo de Fukugita para a amostra apurada (esquerda) e a amostra de trabalho (direita).

A Figura 4.4 mostra a correspondência entre a presente classificação e aquela de Fukugita. À esquerda, a amostra apurada tem um comportamento bastante fiel à realidade. O erro médio é $\langle |T_{normal} - T_{Fukugita}| \rangle = 0.58$, número que aumenta para 1.21 na amostra total. Comparando com a regressão linear, respetivamente 0.71 e 1.07, a distribuição normal permite aperfeiçoar a classificação para as galáxias normais, mas piora a classificação das galáxias ambíguas que foram descartadas da amostra apurada.

Em particular, a classificação baseada em distribuições normais sobre-representa o tipo Sa ($T=2$). Isso se explica pelo fato de que a gaussiana descrevendo a densidade desse tipo é mais larga que as outras. Fora da nuvem de pontos principal formada pela amostra apurada, a densidade de Sa domina grandes volumes do espaço dos parâmetros.

4.2.3 Distribuição empírica

Queremos agora generalizar o método precedente para um distribuição qualquer. Não fazemos nenhuma hipótese sobre uma lei de probabilidade, mas tentamos estabelecer empiricamente qual é a distribuição das variáveis. Para cada tipo morfológico, cons-

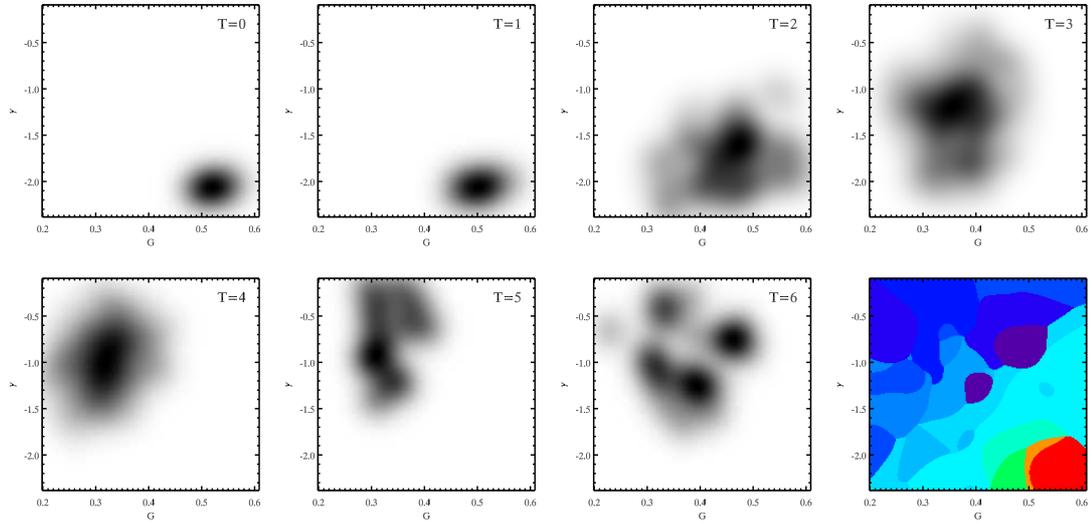


Figura 4.5: Mapas de probabilidade, exemplo no plano G - γ .

truímos um mapa de probabilidade em quatro dimensões que determina qual região do espaço é dominada por qual tipo.

Ao contrário dos outros métodos desse capítulo, este não é analítico. Os mapas de probabilidade devem então ser definidos em cada ponto, discretizando o espaço dos parâmetros¹. Cada galáxia da amostra de treinamento é representada no mapa como um ponto em x . Em seguida, o mapa de pontos é suavizado para espalhar os valores discretos para os pixels vizinhos. A suavização é desempenhada por uma convolução com um *kernel* gaussiano de $\sigma=0.3\sigma_x$, onde σ_x é o vetor dos desvios padrões dos quatro índices morfológicos. Portanto, obtemos uma família de mapas, um por tipo morfológico, definidos sobre todo o espaço dos parâmetros.

Para esclarecer o nosso discurso, mostramos na Figura 4.5 um exemplo de mapas de probabilidade em duas dimensões no plano G - γ . No último mapa, a cor indica o tipo morfológico de probabilidade máxima. Por exemplo, uma galáxia localizada no recanto inferior direito será classificada como elíptica (cor vermelha).

Com esse algoritmo, os mapas de probabilidade podem ser tão afinados que desejado. Basta, para isso, reduzir a largura do *kernel* da suavização. Ao extremo, um

¹Trabalhando com um computador pessoal, o programa IDL não conseguia alocar memória para realizar operações sobre mapas maiores que 40 pixels de lado, ou seja, uma grelha de $40^4=2\,560\,000$ pixels.

4. CLASSIFICAÇÕES MORFOLÓGICAS AUTOMÁTICAS

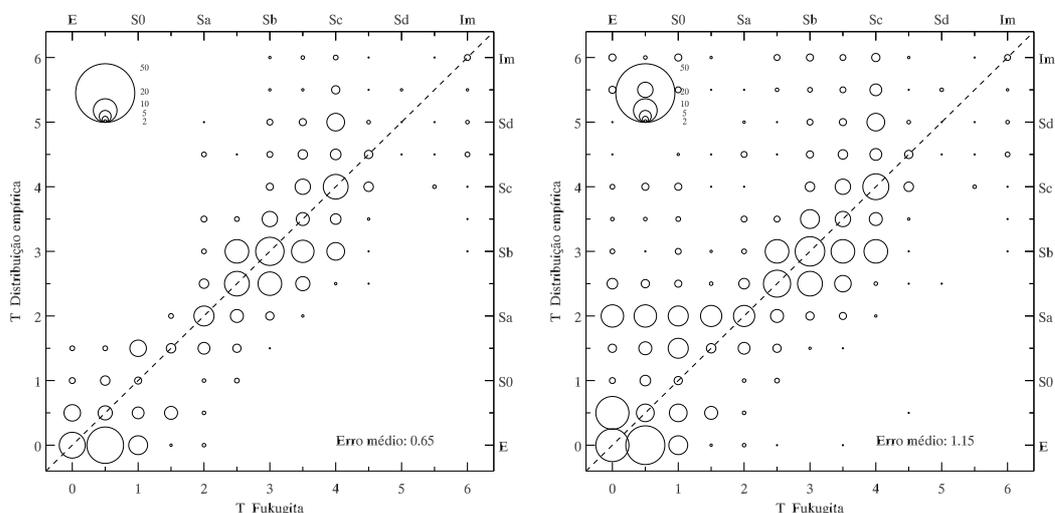


Figura 4.6: Comparação entre o método da distribuição empírica e o tipo de Fukugita para a amostra apurada (esquerda) e a amostra de trabalho (direita).

mapa não suavizado representa exatamente a distribuição da amostra de treinamento, pois foi criado para este fim. Assim sendo, não se pode usar a amostra de treinamento como amostra de teste para avaliar a eficiência da classificação: a taxa de sucesso seria potencialmente de 100%. É necessário separar a amostra em duas subamostras. No entanto, em certos tipos, o número de galáxias se torna perigosamente pequeno para ficar estatisticamente confiável.

A tabela 4.3 mostra o tipo $T_{empírico}$ obtido com a distribuição empiricamente avaliada e as probabilidades de cada tipo.

A Figura 4.6 compara a classificação baseada em distribuições empíricas e aquela de Fukugita. A validade e o comportamento são muito semelhantes à classificação baseada em distribuições normais. Para a amostra apurada, o erro médio é de 0.65, o que é menos bom que com distribuições normais (o erro aumenta). Mas para a amostra total, o erro diminui um pouco de 1.21 para 1.15, o que significa que a classificação empírica é melhor nas galáxias descartadas. Embora atenuado, o problema de sobre-representação do tipo Sa ($T=2$) permanece.

Tabela 4.3: Tipos morfológicos pelo método de uma distribuição empírica.

ID	$T_{Fukugita}$	$T_{empirico}$	$P_{T=0.0}$	$P_{T=0.5}$	$P_{T=1.0}$	$P_{T=1.5}$	$P_{T=2.0}$	$P_{T=2.5}$	$P_{T=3.0}$	$P_{T=3.5}$	$P_{T=4.0}$	$P_{T=4.5}$	$P_{T=5.0}$	$P_{T=5.5}$	$P_{T=6.0}$
Fukugita0001	0.5	1.5	0.12	0.18	0.15	0.30	0.20	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Fukugita0002	3.0	3.0	0.00	0.00	0.00	0.03	0.29	0.27	0.32	0.05	0.04	0.00	0.00	0.00	0.00
Fukugita0003	1.0	0.5	0.24	0.25	0.21	0.20	0.08	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fukugita0004	3.5	4.5	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.17	0.19	0.42	0.08	0.05	0.00
Fukugita0005	3.5	2.5	0.00	0.00	0.00	0.00	0.08	0.65	0.01	0.14	0.12	0.00	0.00	0.00	0.00
Fukugita0006	0.0	0.5	0.34	0.41	0.17	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fukugita0007	4.0	5.0	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.25	0.10	0.17	0.41	0.02	0.00
Fukugita0009	3.0	2.5	0.00	0.00	0.00	0.00	0.36	0.59	0.01	0.03	0.01	0.00	0.00	0.00	0.00
Fukugita0010	2.0	0.5	0.26	0.29	0.22	0.16	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fukugita0012	3.5	3.0	0.00	0.00	0.00	0.00	0.01	0.23	0.46	0.28	0.02	0.00	0.00	0.00	0.00
Fukugita0013	4.0	4.0	0.00	0.00	0.00	0.00	0.00	0.11	0.23	0.16	0.26	0.00	0.23	0.00	0.00
Fukugita0016	1.0	4.5	0.00	0.00	0.00	0.00	0.10	0.02	0.01	0.12	0.16	0.60	0.00	0.00	0.00
Fukugita0017	4.0	3.0	0.00	0.00	0.00	0.00	0.15	0.00	0.85	0.00	0.00	0.00	0.00	0.00	0.00
Fukugita0019	1.0	2.0	0.00	0.01	0.01	0.04	0.40	0.21	0.06	0.27	0.01	0.00	0.00	0.00	0.00
Fukugita0021	0.5	2.5	0.00	0.00	0.00	0.00	0.08	0.90	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Fukugita0022	6.0	4.5	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.16	0.18	0.39	0.11	0.04	0.02
Fukugita0023	1.0	1.5	0.11	0.10	0.22	0.37	0.10	0.06	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Fukugita0024	1.5	0.5	0.26	0.27	0.20	0.19	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fukugita0025	2.5	0.0	0.42	0.25	0.09	0.00	0.01	0.00	0.00	0.17	0.05	0.00	0.00	0.00	0.00
Fukugita0026	6.0	4.5	0.00	0.00	0.00	0.00	0.03	0.03	0.14	0.01	0.01	0.73	0.00	0.00	0.04

4. CLASSIFICAÇÕES MORFOLÓGICAS AUTOMÁTICAS

Esse método de classificação automática, determinando empiricamente a distribuição das variáveis, não tem descrição analítica, requer uma computação fastidiosa e longa, e é mais complexo de aplicar. Apesar disso, não leva melhorias pertinentes de acordo com as classificações precedentes. Para tornar a classificação empírica mais eficiente, os mapas de probabilidades deveriam ser computadas sobre uma amostra de treinamento contendo um número muito maior de galáxias.

4.2.4 Método k-médias

Apresentamos ainda um último algoritmo simples de *clustering*, chamado de k-médias (*k-means*, em inglês). Nesse método, o objeto para classificar é suposto pertencer ao grupo cujo centro é o mais próximo, o que equivale a separar os grupos por hiperplanos.

No espaço dos parâmetros, as galáxias de um mesmo tipo morfológico são localizadas em grupo, cujo centro é, trivialmente, a média $\langle x \rangle_T$ calculada nas Equações 4.4 a 4.16. Para classificar uma galáxia descrita por x , avaliamos sua distância até cada um dos grupos, conforme a norma habitual: $\| x - \langle x \rangle_T \|$. Como os índices morfológicos não têm unidade e são definidos sobre escalas arbitrárias potencialmente muito diferentes umas das outras, devem ser expressos de maneira relativa. O espaço é assim normalizado dividindo os índices por seu desvio padrão. Agora, a classificação k-médias pode atribuir a uma galáxia o tipo cuja distância é mínima. Os tipos obtidos são reproduzidos na Tabela 4.4.

Na Figura 4.7, o tipo determinado pelo método das k-médias é comparado com o tipo de Fukugita. Sumariamente, essa classificação cumpre a sua tarefa com um êxito semelhante às outras. O erro médio sobre a amostra apurada é de 0.62, melhor que os métodos da regressão linear ou da distribuição empírica, mas pior que o método da distribuição normal. Para a amostra total, o erro médio de 1.15 é o mesmo da distribuição empírica, melhor que a distribuição normal, mas pior que a regressão.

Notamos, todavia, que o erro médio é apenas uma indicação da capacidade de uma classificação para reencontrar o tipo de Hubble. De fato, o erro mínimo (0.97) sobre a amostra de trabalho acontece quando se usa uma única regressão. Mas, com esse método, a grande maioria das galáxias são classificadas entre $T=1$ e $T=4.5$; e quase nenhuma é identificada como elíptica ou irregular, o que basta para desacreditar a classificação.

Tabela 4.4: Tipos morfológicos pelas k-médias.

ID	$T_{Fukugita}$	$T_{k-médias}$
Fukugita0001	0.5	1.5
Fukugita0002	3.0	2.5
Fukugita0003	1.0	1.5
Fukugita0004	3.5	4.5
Fukugita0005	3.5	3.0
Fukugita0006	0.0	0.0
Fukugita0007	4.0	4.5
Fukugita0009	3.0	2.5
Fukugita0010	2.0	0.5
Fukugita0012	3.5	3.0
Fukugita0013	4.0	3.5
Fukugita0016	1.0	3.0
Fukugita0017	4.0	2.5
Fukugita0019	1.0	2.0
Fukugita0021	0.5	2.0
Fukugita0022	6.0	5.5
Fukugita0023	1.0	1.5
Fukugita0024	1.5	0.5
Fukugita0025	2.5	6.0
Fukugita0026	6.0	3.5

4. CLASSIFICAÇÕES MORFOLÓGICAS AUTOMÁTICAS

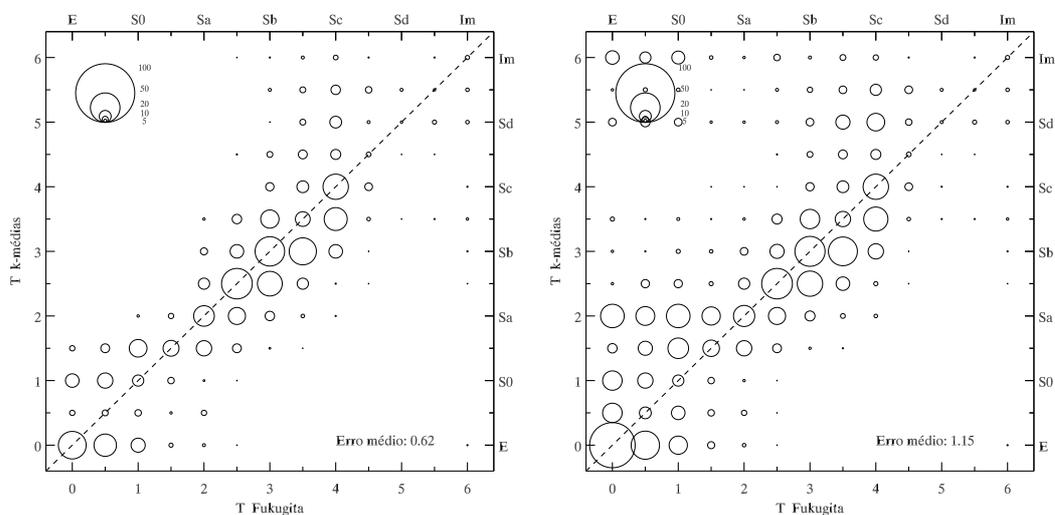


Figura 4.7: Comparação entre a classificação pelo método das k-médias e o tipo de Fukugita para a amostra apurada (esquerda) e a amostra de trabalho (direita).

Concluindo, a eficiência de uma classificação automática depende pouco do algoritmo escolhido. Mais importante é a qualidade das medidas, ou assaz a qualidade dos dados. Para a amostra apurada, todas classificações conseguem prever o tipo de Hubble com um erro médio de 0.6–0.8. O erro cresce para 1.1–1.2 na amostra de trabalho. As galáxias que foram descartadas da amostra de trabalho, porque localizadas longe da média do tipo, apresentam uma ambiguidade morfológica. Estas galáxias comportam um elemento fotométrico que atrapalha as medidas. Uma inspeção visual é capaz de desconsiderar um elemento e se concentrar no resto da galáxia; um algoritmo, não.

Capítulo 5

Anellipticity and granulation as morphological indices to explore star formation in interacting galaxies.

Este artigo foi submetido à revista *Astronomy and Astrophysics*, em 18 de outubro de 2011. Ele retorna a alguns pontos apresentados no Capítulo 3, umas vezes resumindo-os, outras vezes desenvolvendo-os.

Após uma apresentação formal e rápida dos índices morfológicos comumente usados na literatura, C , A , S , G e M_{20} , definimos os nossos dois novos índices, a anelipticidade α e a granulação γ , assim como a isofota de Petrosian. Ele explica como γ é calculada de maneira bastante técnica, usando a transformada de *wavelet* e analisando o seu comportamento sobre galáxias esquematizadas. Ao medir a dispersão radial dos *clumps* de luz, γ pode ser interpretada como a característica que estabelece se uma galáxia contém regiões de formação estelar fora do seu centro.

A estabilidade e a robustez estatística de todos os índices morfológicos são estudadas, avaliando quão forte elas dependem do *redshift*, da nitidez (*sharpness*, em inglês; definida como o tamanho angular da galáxia em unidade de *seeing*), da razão sinal/ruído e da elipticidade. É a nitidez, sobretudo, que deprecia a qualidade e o nível de confiança das medidas de fotometria superficial. Definimos então um critério de qualidade fotométrica como: nitidez ≥ 8 . Estudamos também, prontamente, as correlações entre os índices morfológicos.

5. ANELLIPTICITY AND GRANULATION AS MORPHOLOGICAL INDICES TO EXPLORE STAR FORMATION IN INTERACTING GALAXIES.

Como α mede a deformação da isofota externa, pode ser usado para determinar um “indicador de chance de interação”, $I.C.I. = A \cdot \alpha$. Comparamo-lo com outros indicadores parecidos, aqueles de Conselice (2003) definidos no plano $C-A$ e de Lotz et al. (2004) no plano $M_{20}-G$.

Em seguida, aplicamos esses métodos em catálogos de galáxias HII. Morfologias muito distorcidas são observadas quando as regiões de formação estelar são distribuídas na galáxia e, em média, para galáxias com um grande diâmetro e uma alta taxa de formação estelar específica.

Desses resultados, deduzimos que a formação estelar provocada por uma interação ou fusão não é máxima quando a galáxia apresenta-se muito deformada. Ela se torna máxima em uma fase posterior, quando as instabilidades dinâmicas causam um fluxo de gás significativo para a região central. Durante esta fase final, a formação estelar se localiza essencialmente na região central, mas a morfologia já se normaliza, tornando-se menos irregular. Tudo acontece como se a formação estelar respondesse à morfologia perturbada com um certo atraso.

Anellipticity and granulation as morphological indices to explore star formation in interacting galaxies.

D. Curty¹, E. Telles¹, F. Cuisinier^{†2}, P. Westera³, and C. Rabaça²

¹ Observatório Nacional, Rua José Cristino 77, 20921-400 Rio de Janeiro - RJ, Brazil
e-mail: curty@on.br

² Observatório do Valongo, UFRJ, Ladeira do Pedro Antônio 43, 20080-090 Rio de Janeiro - RJ Brazil

³ Universidade Federal do ABC, Rua Santa Adélia, 166, 09210-170, Santo André - SP, Brazil

Received October 2011

ABSTRACT

Context. The morphology of the galaxies is a key aspect of their evolution. However, the morphological type is not a measurable quantity, but when not assessed by eye, it may be determined from the surface photometry by computing morphological indices. In the era of large data surveys, the acquisition of this information must be entirely automated, what can be improved by mathematical methods of image processing.

Aims. We investigate how the morphological indices are dependent on the image conditions and how they are related to a visually performed classification along the Hubble scheme. We create a quantity that morphologically detects galaxies in interaction and explore how the star forming regions are distributed in interacting galaxies.

Methods. Besides the usual indices, we develop two new morphological parameters: the anellipticity and the granulation. With the first one, we build an indicator of interaction as a measure of distortion of the external isophote; with the second one, we evaluate how the clumps of light are centrally distributed in the galaxy. Then, on two catalogues of HII galaxies, we look how the morphological signature of an interaction relates to star formation.

Results. We determine that concentration-like parameters are more correlated to the Hubble sequence than texture parameters. For HII galaxies, very disturbed morphology is observed when star forming regions are distributed in the whole galaxy and, on average, when galaxies have a large diameter and a high sSFR.

Conclusions. We deduce that the interaction- or merger-driven star formation is not maximum when the galaxy is very distorted. The SFR becomes maximum at a later stage, when dynamical instabilities induce important gas inflow into the central region. At this final stage, the star formation takes place essentially in the central region, but the morphology already becomes less irregular. It seems as if the star formation would respond to the disturbed morphology with a delay.

Key words. Techniques: image processing – Galaxies: structure – Galaxies: star formation – Galaxies: interactions – Galaxies: evolution

1. Introduction

Since Hubble defined his famous sequence, morphological studies were naturally aimed at determining its origin. For this task, it is necessary to firstly measure the Hubble type in a quantitative way. A successful tentative has been done by Sérsic (Sérsic 1963) who fitted the radial surface brightness profile with a power law, as a generalisation of the de Vaucouleurs profile (de Vaucouleurs 1958). To avoid having to assume an underlying model to be fitted, Kent (1985); Conselice (2003) defined the non-parametric concentration index C . The Sérsic index and C are two different numbers to describe the same physical quantity, namely the steepness of the radial light profile. The radial profile is a so specific property of a galaxy that it can be used alone as an approximation of the morphology (Ball et al. 2008; van der Wel 2008).

With the emergence of large data samples, more morphological indices have been proposed. The rotational asymmetry index A (Schade et al. 1995) compares the image of a galaxy to the same image rotated by a half turn. After an idea of Takamiya (1999) and defined by Conselice (2003), the clumpiness index S , sometimes called smoothness, measures the ratio of light com-

ing from clumps to the total light of the galaxy. These three indices form the CAS parameters and lead to a sometimes so-called CAS classification.

More recently, two other indices have been defined. The Gini coefficient G (Gini 1912) is a statistical quantity used to measure inequalities in a population. Adapted to images, it measures the inequality between the pixel intensities and can efficiently be used as a morphological index (Abraham et al. 2003; Lotz et al. 2004). A last index, denoted M_{20} , is the second-order moment of the most luminous pixels containing 20% of the total light of the galaxy (Lotz et al. 2004).

According to the specifically studied galaxies, other indices or parameters have been developed in the literature. By example, Naim et al. (1997) propose for peculiar galaxies four parameters: the blobbiness, the isophotal centre displacement, the filling factor or the middle isophote, and the skeleton ratio. The blobbiness basically aspires to quantify the overall texture just like the clumpiness and the skeleton ratio attempts to evaluate the elongation of detected structures in a residual image.

Yamauchi et al. (2005) propose the coarseness parameter as a measure of the galaxy texture. In concentric elliptical annuli, the deviation of the azimuthal profiles to the isophote is averaged over the radius. Odewahn et al. (2002) does the same job, using

[†] *In memoriam* (1969 – 2011).

Fourier series fits to azimuthal profiles to detect regular spirals or random light peaks.

Interested on irregular dwarfs, Parodi & Binggeli (2003) introduce the lumpiness index and the cluster dimension. The lumpiness index is very close to the clumpiness index, but computed using a median filter instead of a common smoothing. The cluster (or correlation) dimension is the dimension of the fractal structure formed by a discrete set of lumps. This quantity is interpreted as porosity in the intergalactic medium, headquarters of the stochastic self-propagating star formation (SSPSF) (Seiden & Schulman 1990).

A powerful method for morphological studies consists in decomposition of galaxy images. The basis function can be chosen to best meet the considered objects. The shapelets (Refregier 2003; Massey & Refregier 2005; Melchior et al. 2007) are Hermite polynomials weighted by a Gaussian. Kelly & McKay (2004) carry out the polar shapelets decomposition on galaxies and find the coefficients correlating with the Hubble type and other morphological or physical quantities. The Sérsiclets are an hybrid method which combines shapelets and fitting with a Sérsic profile (Ngan et al. 2009; Andrae et al. 2011b). They are used especially in shear measurements of weak gravitational lensing.

Andrae et al. (2011a) give a good overview of different parametrisation schemes and possible pitfalls of certain morphological indices. They point out, for instance, the non-robust measures of the indices, especially C and M_{20} , due to undersampling effects when the computation is performed on too many pixels. Moreover, they warn against the sometimes severe impact of the point spread function (PSF).

Zamojski et al. (2007) calculated the five commonly used indices (C , A , S , G and M_{20}) on large galaxy sample and highlight some very strong correlations. For example, G and M_{20} are correlated and Lotz et al. (2004) argue that a deviation from this broad relation is a signature of merger.

The detection of galaxy interactions or mergers is one of the principal applications of morphological studies. Galaxies in interaction are characterised by tidal forces that create tails and bridges, as shown by Toomre & Toomre (1972) in the first simulation of galaxy-galaxy encounter. Since then, many simulations have confirmed these results and have highlighted the crucial role of gas in interaction/merger-induced starbursts (Barnes & Hernquist 1991, 1996).

In this work, we develop two new morphological indices especially dedicated to interacting galaxies and the star formation dispersion. The first index is the anellipticity α which measures how irregular the outer isophote of a galaxy is. For this purpose, we have come to define the notion of ‘‘Petrosian isophote’’. The second index, based on a multiscale analysis, is called granulation γ and evaluates how the flux is distributed according to the structure sizes. γ is designed to measure how the star forming regions are dispersed in the galaxy and, this way, if the star formation takes place far away from the centre. We use a sample of visually classified galaxies to perform a morphological study and to define an interaction chance indicator. Applying this method to HII galaxies, we show that the star forming regions are spread in the whole galaxy during the beginning of the interaction, when the galaxy shape is very distorted.

This article is structured as follows. Section 2 gives an overview of the the commonly used morphological indices. In Section 3, our three working samples are presented. In Section 4, we define two new morphological indices and briefly study the robustness of the measurements and the correlations between them. Section 5 is dedicated to the definition of an interaction

chance indicator. With help of the latter, we discuss in Section 6 the star formation distribution of the HII galaxies.

2. Commonly used morphological indices

The concentration index C is computed from the curve of growth of the galaxy intensity profile by relating two galactic radii containing a certain fraction of the light flux. We applied the definition of Conselice (2003), but by using elliptical apertures to improve the measurement as shown in Doi et al. (1993); Yamauchi et al. (2005).

$$C = 5 \log \left(\frac{a_{80}}{a_{20}} \right) \quad (1)$$

where a_{80} and a_{20} are the semi major-axes of the ellipses containing 80% resp. 20% of the total flux.

The asymmetry index A is the fraction of the flux that has no symmetric counterpart. It is computed by building a residual image, which is the difference between the original image I_0 and I_{180} , the image rotated by 180 degrees. The asymmetry A is then the fraction of the flux in the residual divided by the total flux. Following the method of Conselice et al. (2000), the rotation centre is chosen by an iterative process, to minimise A . We adopt the following definition:

$$A = \frac{\sum |I_0 - I_{180}|}{\sum I_0} - A_{sky}. \quad (2)$$

The contribution of the sky noise A_{sky} is handled in different ways in the literature. Conselice et al. (2000) explicitly compute the asymmetry index in an empty region of the sky. Zamojski et al. (2007) estimate A_{sky} by calculating a second asymmetry on a slightly convolved image. They assume A_{sky} to decrease proportionally to the size of the convolution kernel and are, this way, able to evaluate the effect of the noise on A . On our side, we estimate theoretically the sky contribution for a Gaussian distribution of the noise. If the standard deviation from the background, measured in an empty part of the sky, is σ_{sky} , the noise in the residual would be $\sqrt{2} \sigma_{sky}$. The mean flux per pixel for a half-normal distribution can easily be calculated and the asymmetry of the sky is then

$$A_{sky} = \frac{2N\sigma_{sky}}{\sqrt{\pi}} \frac{1}{\sum I_0}, \quad (3)$$

which can be replaced in 2.

The clumpiness index S is the fraction of light of the galaxy contained in clumps. The clumps are determined by subtracting a smoothed image I_{smo} from the original one I_0 . The smoothing kernel is a box whose side length is proportional to the Petrosian radius and chosen at $1/4r_{Petro}$. To prevent that the galactic centre would be considered a clump, which would happen with the very cuspy de Vaucouleurs profile, the centre is blanked. We chose for this a mask $1/4r_{Petro}$ in size. Once again, the contribution of the background is treated differently among the authors. To avoid negative values of S , we consider as clumps only the residual pixels brighter than a certain threshold, fixed at 3σ of the noise. Thus, we adopt the following definition of the clumpiness:

$$S = 10 \frac{\sum_{pix > 3\sigma} (I_0 - I_{smo})}{\sum I_0}. \quad (4)$$

The Gini coefficient G measures the inequality of the pixels’ intensities. It is based on the Lorenz curve (Lorenz 1905)

which describes, in economics, the fraction of the total wealth belonging to the poorest members of a population. To compute G , the pixels are first sorted from the lowest to the highest intensity to form the function $I(y)$, where y is the index of the pixel. The cumulative sum of this function is the Lorenz curve $L(x)$ in function of the pixel index x :

$$L(x) = \sum_{y < x} I(y) \quad (5)$$

The Lorenz curve is a way to describe the distribution of the flux in the pixels. When normalised by the total flux, $L(x)$ measures the fraction x of the pixels containing the fraction L of the total flux. For example, if all pixels except one have intensity zero, which is complete inequality, the Lorenz curve is zero everywhere except for the last point; in this case, a fraction $x \neq 1$ of the pixels contains a fraction $L = 0$ of the total flux. If all the pixels have the same value, which is perfect equality, the Lorenz curve is a straight line with positive slope; in this case, a fraction x of the pixels contains a fraction $L = x$ of the total flux.

The Gini coefficient is defined as twice the area between the Lorenz curve and the diagonal connecting the first and the last point of the Lorenz curve, i.e.:

$$G = 1 - 2 \frac{\sum L}{N_{pix}}. \quad (6)$$

G can take on values between 0 (perfect equality) and 1 (complete inequality). By calculating G with different apertures, Lisker et al. (2006) are able to detect bars in early-type galaxies. Consequently, Lisker (2008) demonstrated the crucial importance of the aperture in the computation of G . More precisely, he warns against a too large aperture of $1.5 R_{Petro}$, frequently used in the literature, which leads to values of G completely biased by the signal-to-noise ratio (S/N). To avert this annoying effect, he recommends using an aperture of $1 R_{Petro}$ or, even better, $2/3 R_{Petro}$. For calculating the Gini coefficient, we reduce the aperture to $2/3 R_{Petro}$.

The M_{20} index is the second-order moment of the most luminous pixels. To establish, which are the pixels containing 20% of the total light of the galaxy, one must use the Lorenz curve again. After determining, which pixels have to be summed up, one calculates the central second-order moments to get M_{20}

$$M_{20} = \log \left(\frac{\sum_{20\% \max} I r^2}{\sum I r^2} \right), \quad (7)$$

where r is the distance of the pixel to the centre.

The second-order moment is a measure of the dispersion around the centre. If the 20% most luminous pixels are all in the bulge, M_{20} remains small and then correlates with C (Lotz et al. 2004; Zamojski et al. 2007) for normal galaxies.

3. Work samples

3.1. Morphological catalogue

For the morphological study, we use the galaxy catalogue of Fukugita et al. (2007), derived from the Sloan Digital Sky Survey – SDSS (York et al. 2000). This sample contains all the galaxies in a rectangular region of the equatorial area in the northern sky with an r -magnitude < 16 . Among the total of the 2253 galaxies, 1866 have spectroscopic information. The mean

redshift is $z_{mean} = 0.056$ and the redshifts do not exceed $z_{max} = 0.166$. The galaxies, of all morphological types, were visually classified by the authors using a single index T along the Hubble sequence. This Hubble index runs from $T = 0$ for ellipticals, $T = 1$ for lenticulars, $T = 2 - 5$ for spirals from Sa to Sd, until $T = 6$ for Magellanic irregulars, allowing half-integer values for intermediate types. The value $T = -1$ is assigned to 35 galaxies which cannot be classified by a conventional Hubble type. Those galaxies are often amorphous, sometimes clearly in interaction. In their catalogue, Fukugita et al. (2007) identify 30 interacting galaxies.

The SDSS photometric pipeline (Lupton et al. 2001) tends to over-deblend galaxies with angular sizes over 2 arcmin (Kushner et al. 2006). For instance, *Fukugita0048* and *Fukugita0049* are two parts of the same spiral galaxy. To avoid such problems, we chose to work out the original images of SDSS and to develop our own simple cleaning procedure. The galaxies are first extracted from the field images. We cut a square area at $3r_{Petro}$ around the galaxy centre. Then, an algorithm attempts to erase field stars or neighbour galaxies: (1) A first approximation to the galaxy shape is made with a histogram based thresholding; (2) The objects outside this shape are detected by histogram thresholding; (3) A morphologic dilation operator is implemented on the area of the detected objects (the size of the structuring element is proportional to the size of the area); (4) The detected objects replaced by random pixels following the Poisson distribution of the sky; (5) Eventual field stars inside the galaxy's shape are not removed. The algorithm is applied simultaneously in the 5 photometric bands of the SDSS, *ugriz* and is repeated three times with refined thresholds in each loop.

As the visual classification was carried out on the g -band images by Fukugita et al. (2007), we computed the morphological indices in this same colour.

3.2. HII galaxy catalogues

As we want to concentrate on galaxies with strong star formation, we chose two HII galaxy catalogues. The first one is the SDSS HII galaxies with Oxygen abundances Catalogue – SHOC (Kniazev et al. 2004), which contains 624 spectra belonging to 612 galaxies with a mean redshift of $z_{mean} = 0.066$ and a maximum of $z_{max} = 0.358$. The second HII galaxy catalogue is a sub-sample of the SDSS too, defined by Izotov et al. (2006). It contains 309 spectra with a mean redshift of $z_{mean} = 0.028$ and a maximum of $z_{max} = 0.239$.

4. Two new morphological indices

We wish to quantify the morphological disturbances and star formation distribution of galaxies. We define for this purpose two new parameters: the anellipticity index α and the granulation index γ . While the first evaluates the deviation of the outer isophote from the ellipse, the second measures how the galaxy flux is spread according to the structure sizes.

4.1. The anellipticity index α

The idea of using the distortion of the outer shape of the galaxy as an indicator of interaction or merger is old, both in ellipticals with boxy isophotes (Lauer 1985) and in very late types (Naim et al. 1997; Telles et al. 1997). The anellipticity index α attempts to perform this task automatically in a formal way and without assumption of symmetry as in the case of boxy isophotes. γ

quantifies the irregularity by comparing the galaxy outer shape to an ellipse.

The choice of a suitable outer limit of a galaxy is a key problem in surface photometry. A too small aperture could bias the results, because only the central part of the galaxy would be considered, whereas in too large aperture, the sky noise would contaminate the measurements. To ensure taking into account the whole galaxy, as independently as possible of the observation conditions, the segregation map must be defined without resorting to the sky noise. The outer limits of the objects have to be specified intrinsically, as is the case for the Petrosian radius (Petrosian 1976). Since we want to conserve the general aspect of the outer shape of the galaxy, we want to use an isophote and define therefore the ‘‘Petrosian isophote’’.

The Petrosian radius is the distance r_{Petro} where the local surface brightness μ is a given fraction, generally chosen at 0.2, of the mean surface brightness within r_{Petro} :

$$\frac{\langle \mu(r = r_{Petro}) \rangle}{\langle \mu(r < r_{Petro}) \rangle} = 0.2 \quad (8)$$

Similarly, we define the Petrosian isophote as the isophote where the surface brightness is 0.2 times the mean surface brightness of the area it encloses. As the intensity of the Petrosian isophote is close to the sky level, it is often distorted by background noise. To get regular, not too noisy isophotes, a threshold is applied to a filtered image and the so obtained mask itself is once again smoothed. Note that our Petrosian isophote is only slightly different from the segmentation map used by Lotz et al. (2004). Instead, they use the isophote of surface brightness $\mu = \langle \mu(r_{Petro}) \rangle$.

In Fig. 1, three Petrosian-like apertures are shown for the irregular galaxy *Fukugita2244*. The first image, on the left, is a circular aperture with the common Petrosian radius. We can see that a quite large part of the sky is contained in the area assigned to the galaxy. The large number of noisy pixels within this area plays an important role in the calculation of the Gini coefficient so this aperture is definitively not the best choice. The second image, in the middle, is the Petrosian ellipse, i.e. the Petrosian ‘‘radius’’ computed with elliptical aperture. The image on the right shows our Petrosian isophote. All the morphological and photometric quantities are computed inside the Petrosian isophote, except for the Gini coefficient. For the latter, we shrink the mask by a factor of 2/3, as suggested by Lisker (2008).

In order to measure the irregularity of the isophote, we create a dimensionless parameter by dividing the square of its perimeter P by the area A within¹. This number is dominated by the length to width ratio of the galaxy shape. To break the dependence on the ellipticity, we calibrate our parameter using the ellipse of maximum and minimum spatial root mean square, which is computed from the second-order moments (see, for example, Bertin & Arnouts (1996) and the SExtractor manual). So, our parameter, denoted α , measures the deviation of the galaxy shape from the ellipse. That is why we call it the anellipticity index.

$$\alpha = \frac{P^2 A_{ell}}{A P_{ell}^2} - 1 \quad (9)$$

We subtract unity in analogy to the definition of the ellipticity and to force α to take on values close to 0 for regular isophotes.

Note that our anellipticity index is somehow similar to the ‘‘filling factor’’ defined by Naim et al. (1997). Their parameter

¹ The inverse ratio is used in the isoperimetric quotient $Q = 4\pi A/P^2$.

is the ratio of the area of the isophote to the area of the circumscribed ellipse. Nevertheless, they apply the measurement on a middle isophote and assume the ellipse to have same position angle and ellipticity as the whole galaxy.

For a given pair of semi-major and -minor axes a and b , the area of the ellipse can be expressed analytically: $A_{ell} = \pi ab$. But there is no exact analytical formula for the perimeter. Therefore, we use Ramanujan’s second approximation:

$$P_{ell} = \pi(a+b) \left(1 + \frac{k}{10 + \sqrt{4-k}} \right) \quad (10)$$

where

$$k = 3 \left(\frac{a-b}{a+b} \right)^2 \quad (11)$$

Basically, α is close to zero for normal galaxies. Only very irregular shapes, such as the ones due to the presence of tails, bridges or a companion galaxy, can lead to significant non-zero values of α . It is worth noting that α can be negative, because the ellipse is not the smallest perimeter for a given area and ellipticity. For a high length to width ratio, like in edge-on galaxies, a boxy isophote has a greater interior area than the corresponding ellipse and α becomes negative, albeit close to zero. On the contrary, an ‘‘X’’-shaped isophote leads to positive values α . Despite this, our anellipticity index does not show a relation with the 3rd and 4th order moments as calculated by the IRAF task *ellipse*. The 4th moment effectively describes the boxiness only if the galaxy is quite symmetric, when the other orders of the circular harmonic decomposition have small amplitudes. Moreover, IRAF’s *ellipse* stops computing when the isophote reaches the sky noise, which does not allow a rigorous comparison.

4.2. The granulation index γ

In a galaxy, clumps of light are principally formed by the ionised gas of HII regions, the ionisation being caused by young, very massive stars. We seek to measure how the star forming regions are distributed in the galaxy. More precisely, we want to compare the intensity and the dispersion of the HII regions to the ones of the older stars.

The HII regions are small compared to the size of the whole galaxy. We assimilate small scale structures to star forming regions and large scale structures, as the disc, to population II stars. If we were able to separate a galaxy into different structures according to their sizes, we could get photometric information about the star formation dispersion. We define for this purpose the granulation index γ , based on a multi-scale method: the wavelet transform (WT).

Similar to the Fourier transform, the WT is a general method of data analysis. We use in this work the OV-WAV package (Pereira & Rabaça 2003): written in IDL, it resorts to the *à trous* algorithm and is designed for bidimensional data. Previous studies have already been performed with this package, including the analysis of planetary nebulae structures (Cuisinier et al. 2005; Leal Ferreira et al. 2009); the detection, in groups of galaxies, of tidal debris as small scale structures (Epitácio Pereira et al. 2007) or diffuse light as large scale structures (Da Rocha & Mendes de Oliveira 2005); and the morphological classification of galaxies (Curty et al. 2010). The Appendix A gives a brief overview of the WT and the *à trous* algorithm.

The WT extracts from the original image only the components of a certain size. We want to assess how the galactic flux is

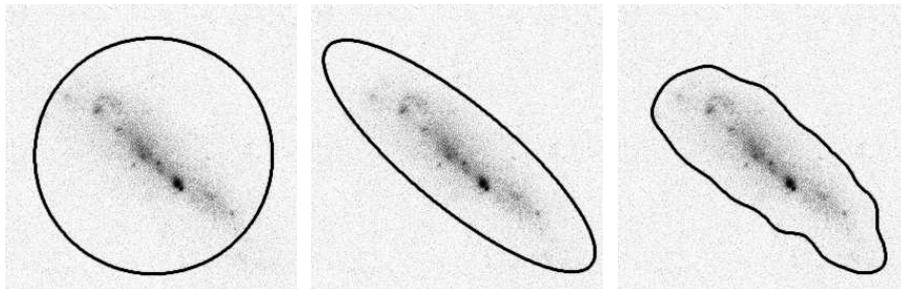


Fig. 1. The irregular galaxy *Fukugita2244* surrounded by 3 different Petrosian-like apertures. On the left, the common Petrosian radius computed with circular aperture. In the middle, the axes of the Petrosian ellipse are determined with elliptical apertures. On the right, the Petrosian isophote as we defined it.

spread according to the scale of the WT. A very late-type galaxy, with bright HII regions, contains many small scale structures, while an early-type does not. Unfortunately, the cuspy centers of elliptical galaxies result in important fluxes in the small scale WT coefficients. For this reason, it is not convenient to analyse directly the raw wavelet spectrum.

With the *à trous* algorithm, the WT coefficients c_i are a set of images with the same dimensions as the input image. It is then possible to handle them to compute quantities more relevant than just the amplitudes given by a Fourier analysis. This approach is made feasible because the WT conserves the spatial information.

To reduce the influence of the central flux and give more emphasis to the flux in the outer regions, we weigh the pixels intensity with their squared distance to the galactic centre. This is, by definition, the central second-order moment. To avoid projection effects, we use elliptical apertures and compute the second order moment using the the deprojected radius r_c . The weighted spectrum A_i is a function of the octave i and is defined by

$$A_i = \sum c_i r_c^2. \quad (12)$$

In addition to being modified by inserting weights, note that A_i is not a power spectrum but an amplitude spectrum, because we do not square the coefficients. Eventual negative fluxes and the noise contribution are discarded by defining a zone of interest.

If normalised by the total flux, the central second-order moment is the variance σ^2 , i.e. a measure of the dispersion around the centre. As we do not normalise, this way we evaluate, in each of the wavelet coefficients, both the flux F_i and the dispersion of the flux σ_i^2 . The weighted spectrum can be written as

$$A_i = F_i \sigma_i^2. \quad (13)$$

We now define the granulation index γ as the slope of this weighted spectrum A_i in function of the scale. To reproduce the large intensity variation and turn γ more intuitive, the amplitude spectrum is expressed in a decreasing logarithmic scale, like a magnitude. The inclination is determined through standard least squares linear regression. The first point is excluded from the fit because it corresponds to a structure size of one pixel, incompatible with the seeing and completely overpowered by the noise. The scales larger than the Petrosian diameter are excluded from the fit too.

In summary, the calculation of the granulation index γ involves the following steps:

1. Compute the WT on the image;

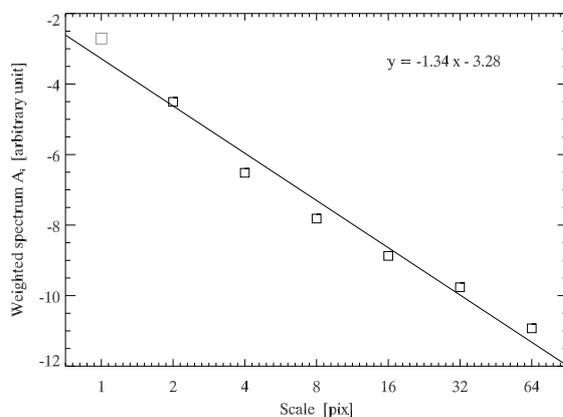


Fig. 2. Wavelet spectrum of the galaxy *Fukugita2244* (squares). The line is the result of a linear regression. The first point is excluded from the fit.

2. Compute the central second-order moment of each WT coefficient using elliptical aperture, which leads to weighted spectrum A_i ;
3. Express the wavelet spectrum as magnitudes;
4. Fit a standard linear regression to the spectrum;
5. The granulation index γ is the inclination of the fitted line.

Figure 2 is the amplitude spectrum of *Fukugita2244* resulting from the wavelet transform shown in Fig. A.2 of Appendix A. The slope of the regression line is the granulation index: $\gamma = -1.34$.

The spectrum A_i measures the flux and its dispersion coming from structures with sizes of 2^i pixels. Large scale structures, typically formed by the disc, generally contribute most of the flux and are, in all cases, more widely distributed. The actual spectrum is increasing, but the negative sign in the definition of the magnitude causes A_i to decrease.

Interpreting Equation 13, and keeping in mind the fact that γ is the inclination of A_i , one can already forecast the behaviour of γ . For elliptical galaxies, the only small scale structure is the centre, which contains much of the flux. But as it is less spread out than the large scale structures, A_i is very steeply inclined and γ becomes very negative. At the other end of the Hubble sequence, very late type galaxies emit an important part of their flux in small scales structures and these small clumps are

distributed within the whole galaxy, so that A_i is much flatter and γ close to zero.

But what is the physical meaning of γ , and how does it differ from the clumpiness index S ? To answer these questions, we create artificial images of schematic galaxies. They are formed by a disc and some luminous peaks as clumps. We seek to clarify how γ responds when we modify the size, intensity, radial distribution and number of the clumps.

A first family of three galaxies is composed of 100 clumps normally distributed around the centre, as can be seen in Fig. 3. To simplify, both the disc and the clumps are circular Gaussians. The full width at half maximum (FWHM) of the clumps are 0.05 (1st column), 0.1 (2nd column) and 0.2 (3rd column) times the FWHM of the disc. The distribution of the clumps in the disc follows the flux distribution of the disc, that means the density of the clumps is a Gaussian function with the same FWHM as the disc. The total flux emitted from the clumps equals the flux from the disc, i.e. half the total flux of the galaxy.

In a second family of three galaxies, we reduce the flux of the clumps to attain 1/10 of the total flux. The next three galaxies are still composed of 100 clumps but their distribution is tighter around the centre: it has half the FWHM of the disc. In the fourth family, the last row of Fig. 3, the galaxies have only 20 clumps containing half the flux of the galaxy. This means that a clump in the last row is 5 times more luminous than a clump in the first row. The calculated values of C , S and γ are displayed at the bottom of the images.

The disc represents the old population consisting of cold and red stars. The clumps can be associated to HII regions induced by the very young population of hot and blue stars. Thereby, the flux fraction in the clumps is basically proportional to the specific star formation rate (sSFR). A narrower distribution of the clumps models a compact galaxy with a high central star formation rate. Most of the blue compact dwarfs (BCDs), for instance, show a significant fraction of gas migrating to the centre (Hunter & Elmegreen 2004). Models with fewer but brighter clumps reproduce galaxies with star formation taking place in giant clusters. For a given SFR, this property implies a smaller cluster formation rate (CFR) and a higher fraction $\Gamma = \text{CFR}/\text{SFR}$ (Bastian 2008; Goddard et al. 2010).

Figure 4 shows the morphology parameters of these model galaxies in the $C - \gamma$ plane at the top and in the $S - \gamma$ plane at the bottom. On the left, the 2253 galaxies of the morphological sample are plotted with a colour code indicating the Hubble type as evaluated visually by Fukugita et al. (2007). On the right, the green squares, blue triangles, purple diamonds and red asterisks are the simulated galaxies of the 1st, 2nd, 3rd and 4th rows of Fig. 3, respectively.

It can be seen in the two left panels of Fig. 4 that the vast majority of the elliptical galaxies have nearly the same value of γ , whereas late type galaxies have higher γ values. This bimodal distribution of γ shows the high capacity of this index to separate early type from late type galaxies. For late type galaxies, γ correlates with the logarithm of S (bottom left panel). Indeed, these two indices are related to the flux fraction in the clumps. For the model galaxies, as expected, C remains constant when varying only the size of the clumps; and when the clump density becomes more concentrated, C increases as well (upper right panel of Fig. 4).

For the model galaxies, increasing the clump size leads to a simultaneous decrease of S and γ , parallel to the slope of the correlation of the Fukugita galaxies. More interestingly, the green squares and the red asterisks have the same γ values: the granulation index is not sensitive to the number of clumps. Also, the blue

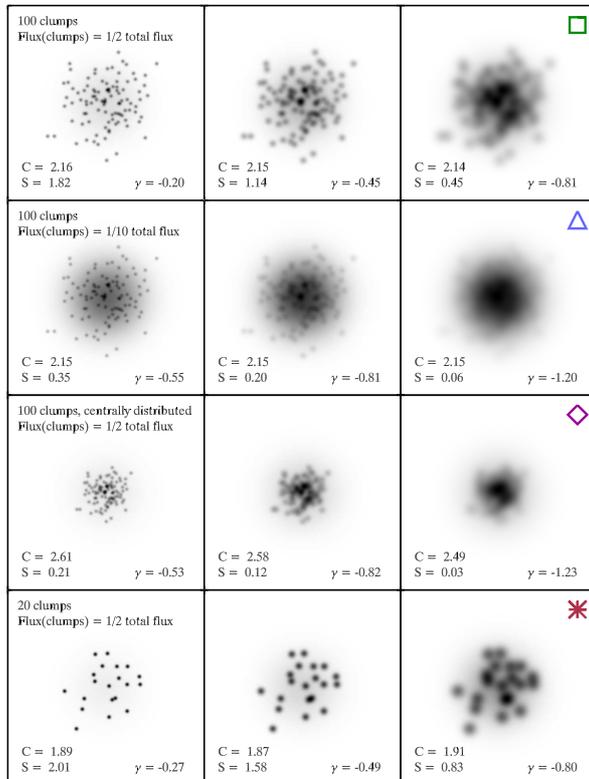


Fig. 3. Images of schematic model galaxies. In the 1st row, the three galaxies are composed of a disc and 100 clumps. In the 2nd row, the clumps are less luminous; in the 3rd row, the clumps are more centralised. In the 4th row, the clumps are less numerous. The radial profiles of the clumps are 0.05, 0.1 and 0.2 times the radial profile of the disc in the 1st, 2nd and 3rd columns, respectively. The images are grey-scaled logarithmically. The symbols to the upper right of the last panels of each row are the ones used in Fig. 4.

triangles and the purple diamonds have the same γ : the intensity of the clumps and their spatial spreading have effects that cancel each other. As the weighted spectrum defined in Equation 13 is the product of intensity by the radial distance, decreasing the intensity of the clumps and concentrating them closer to the centre have similar aftermath on γ .

From this simulation, it appears that S and γ partially measure the same thing, that is the fraction of the flux contained in the clumps. The squares, diamonds and asterisks represent galaxies that have all half of the flux in the clumps. The initial purpose of the clumpiness index was to measure exactly this fraction. So ideally, S should have the same value for the galaxies represented by these three symbols, independently of the clump size. This is clearly far from the case, because it is not possible to recover every clump in an image, with blended clumps and discs. However, we note that our granulation index responds more strongly to the clumps of small scales than S . Also, γ is more sensitive to clumps located far away from the centre, but does not care about their number. This way, γ evaluates the intensity of the star formation far away from the centre.

To visualise this with real data, we pick 10 very late-type (Sc, Sd, Im or Unclassified) galaxies from the morphological catalogue. The galaxies, chosen to follow more or less vertical

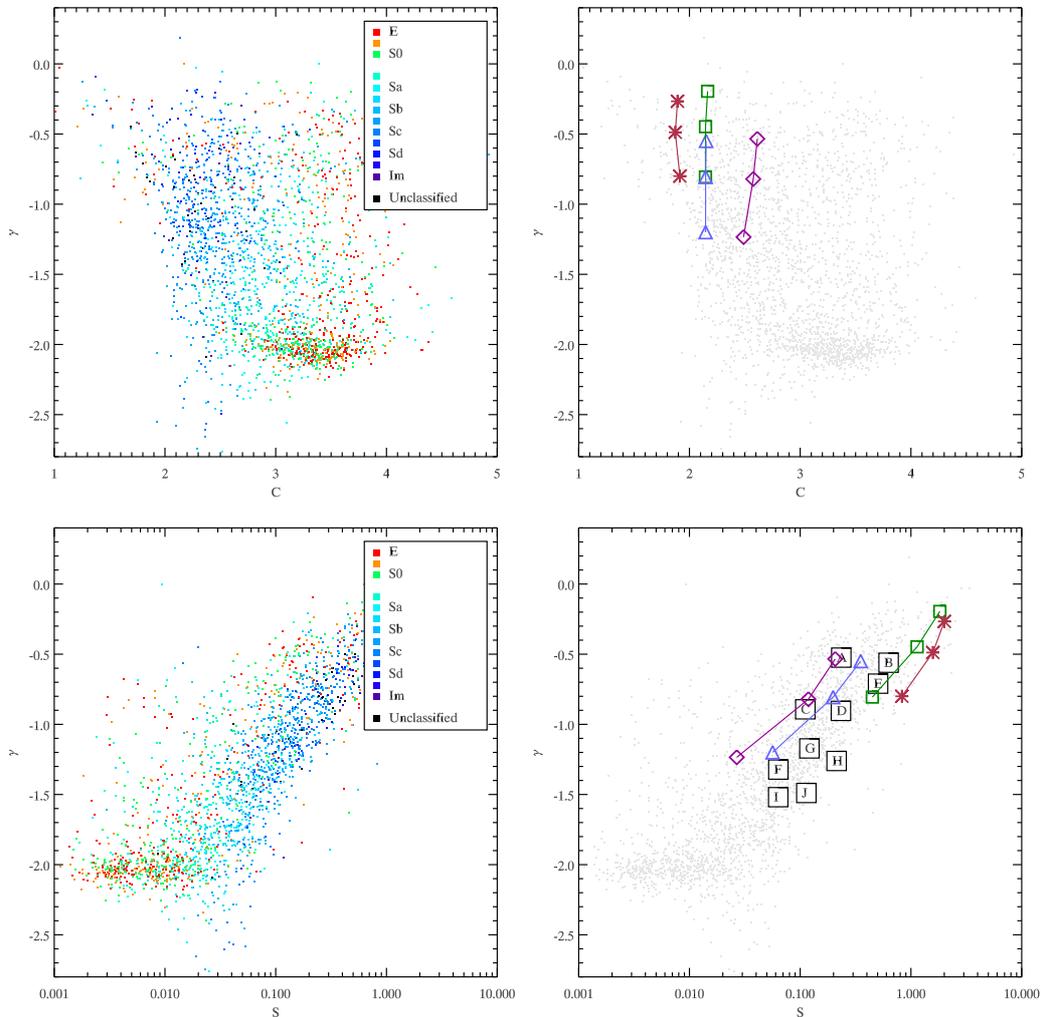


Fig. 4. On the left, the C - γ (top) and S - γ (bottom) planes for the 2253 galaxies of the morphological sample colour coded by morphological type. On the right, the same diagrammes for the schematic galaxies: the green squares, blue triangles, purple diamonds and red asterisks represents the four rows of Fig. 3. The letters from A to J correspond to the late type galaxies shown in Fig. 5.

and horizontal lines in the S - γ plane, are tagged from A to J in Fig. 4. Figure 5 shows them with their relative position like in Fig. 4, lower panels.

4.3. Robustness of the morphological indices

Image processing provides powerful tools for surface photometry analysis, but all the measurements depend upon the quality of the image we are handling. The quality of the data is affected not only by instrumental or observational conditions, such as the seeing, the sky noise or the angular resolution, but also by redshift or projection effects. Ideally, a robust methodology should not depend on any of these parameters.

For instance, at low S/N , the Petrosian isophote can be very close to the sky level and the isophote, despite the applied smoothing, could remains noisy or jagged, which greatly increases the anellipticity index. This happens for elliptical galax-

ies because the de Vaucouleurs profile tends slowly to zero and the curve of growth consequently increases steeper even far away from the centre. More frequently, a field star or a neighbouring galaxy has not been properly deleted by the cleaning process. If it is close enough to enter in the Petrosian isophote, it will be considered a part of the galaxy! This explains why certain galaxies, even classified as ellipticals, take on high α values without reflecting an interaction signature.

Two examples of bad cleaning affecting the Petrosian isophote are shown in Fig. 6. In the upper right panel, the isophote makes a detour to enclose a still present background galaxy on the left side. The rest of the isophote is a bit irregular because its magnitude is close to the sky noise. In the bottom right, the bright star contributes significantly to the inner flux during the computation of the Petrosian isophote. This leads to a too high and incorrect Petrosian isophote level, such that the isophote does not enclose the spiral arms. These examples illus-

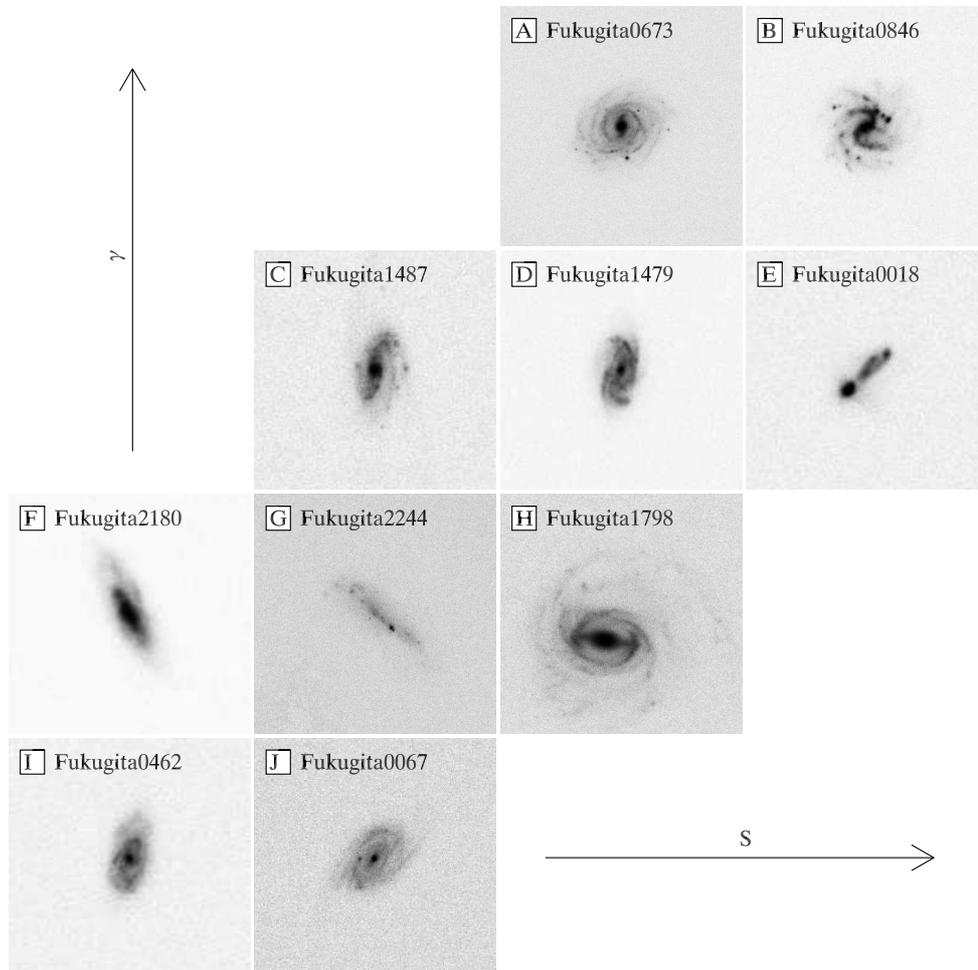


Fig. 5. S - γ plane with the 10 galaxies tagged from A to J in Fig. 4. The images are in the g -band, resized, and drawn in logarithmic scale.

trate how the surface photometry can be greatly misinterpreted during a completely automated process.

To ensure the robustness of the morphological indices, we seek behaviour patterns in function of observational conditions. A reliable measurement should be as little as possible dependent on the redshift, the resolution, the apparent ellipticity or the S/N ratio. The S/N – actually the S/N per pixel – is defined as the average flux of the galaxy per pixel, divided by the sky noise σ_{sky} . We denote as “sharpness” the angular size of the galaxy in unit of seeing, i.e. the dimensionless parameter obtained by dividing the angular diameter by the full width at half maximum (FWHM) of the PSF. We prefer this quantity to the mere seeing because a very well resolved galaxy would not be affected by bad seeing. To test the stability of the indices, we separate the morphological catalogue into quartiles according to the four quantities: redshift, sharpness, S/N ratio and ellipticity. For a given morphological type, the value of a morphological index should not vary much in function of the observational conditions. In every quartile and for every morphological type, we compute the median and the mean absolute deviation of the 7 morphological indices. In order not to clutter Figure 7, we divide the sample in only four

groups: E/S0 (red diamonds), Sa/Sb (light-blue squares), Sc/Sd (dark-blue triangles) and Im and unclassified (purple crosses).

In Fig. 7, one can see that even the widely used concentration index is not very robust: it depends quite heavily on the sharpness, on the S/N ratio and, although working with elliptical apertures, on the ellipticity too. On average, α remains close to zero for all morphological types. It becomes unstable for low S/N , due to the difficulty to establish the Petrosian isophote in this case. In a particular way, γ takes on high values for a large sharpness, when the galaxies are very well resolved. This is simply the influence of a field star or a background galaxy, which is frequent when the galaxy is seen with a large angular size. The presence of a foreign object close to the outer limit of the galaxy makes γ blow up. The ellipticity has an influence on indices which require smoothing. It is noticeable, despite the complex methodology based on a multiscale analysis, that γ stays very stable under all the observational conditions. On average, α is very close to zero, showing that the outer isophote is nearly always elliptical. Only some irregulars present very distorted outer isophotes, which must be interpreted as an interaction signature.

As a consequence of this robustness study, for all the coming calculations, we dismiss the galaxies with an angular diam-

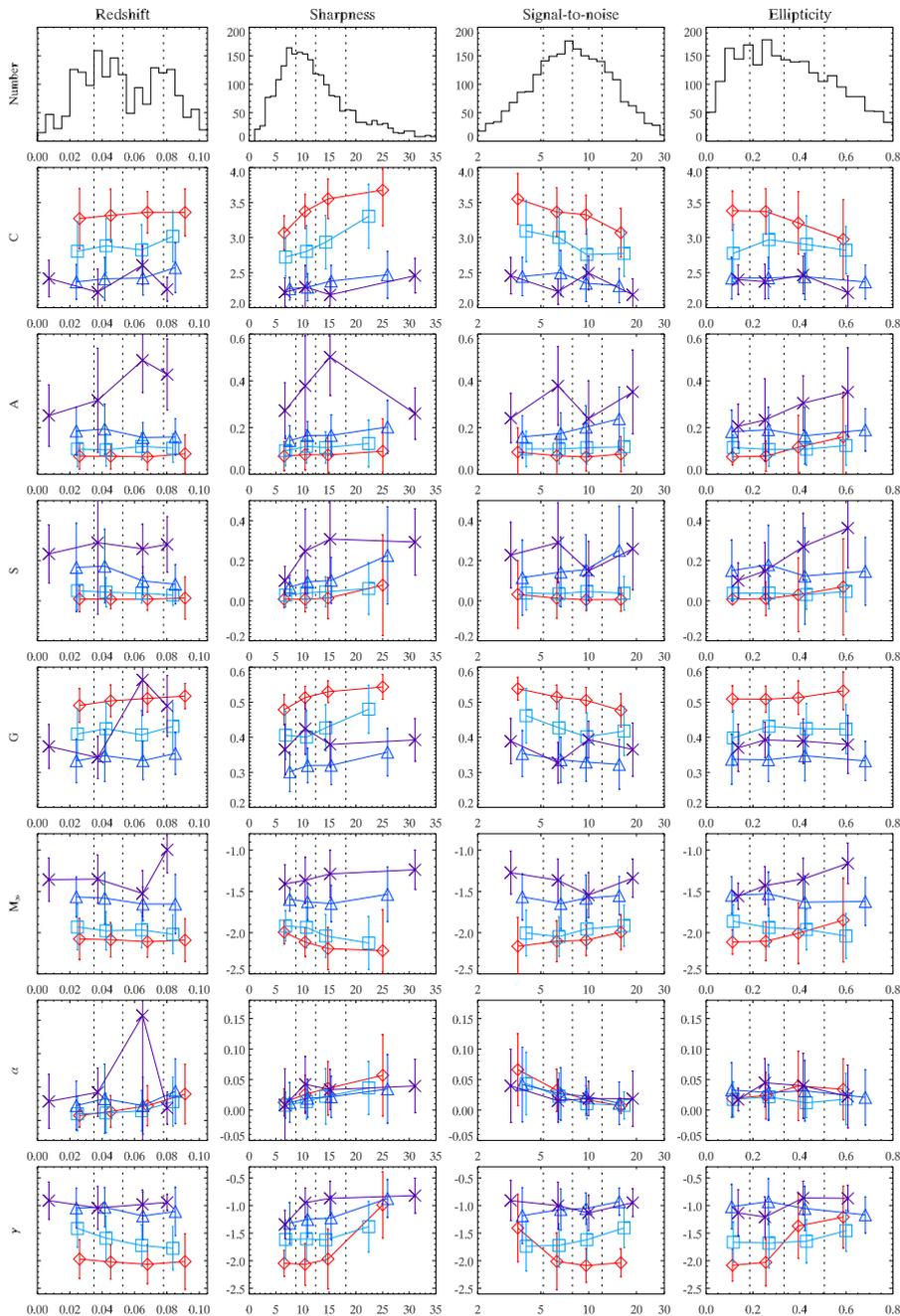


Fig. 7. Robustness of the morphological indices. The morphological sample is separated in four groups: E/S0 plotted as red diamonds, Sa/Sb as light-blue squares, Sc/d as dark-blue triangles, and finally Im and unclassified as purple crosses. These four groups are separated in quartiles according to, from the first to the last column: redshift, sharpness (=angular diameter divided by the seeing), S/N ratio and ellipticity. In the quartiles, we plotted the median of the morphological indices and the mean absolute deviation as error bars.

eter too small compared to the seeing. Thus, we only consider galaxies with sharpness ≥ 8 . Although this criterion may seem quite permissive for the morphological sample with 80% of the galaxies fulfilling it, it is rather restrictive for the HII galaxy sam-

ples: only 294 galaxies (31%) satisfy the condition. Besides being on average more distant, the galaxies of the Kniazev and Izotov catalogues are mostly dwarfs. The large proportion of rejected galaxies could lead to the conclusion that the resolution

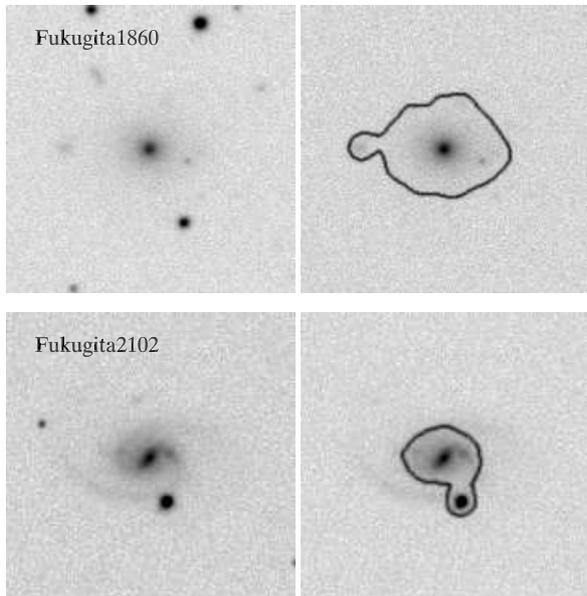


Fig. 6. An elliptical galaxy (top) and a spiral galaxy (bottom) before and after the cleaning process. On the left, the original field images show foreground stars and background galaxies. On the right, the same images after cleaning still contain objects foreign to the galaxies which affect the Petrosian isophotes (thick lines).

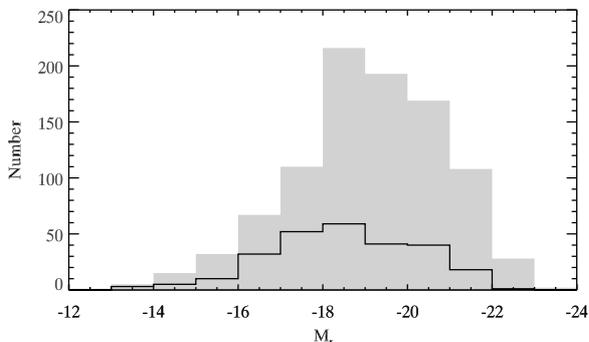


Fig. 8. Distribution of absolute r magnitudes for all HII galaxies (grey) and for the ones satisfying the photometric resolution criterion (solid line).

criterion discards all the compact dwarfs. To ensure not to introduce a selection bias, we compare the distributions of absolute r magnitudes. We plot in Fig. 8 the histograms for all the HII galaxies and for the well resolved ones. It can be seen that the general shape of the histogram is the same. The quality criterion does not eliminate especially dwarfs, but conserves every magnitude in similar proportion. This proves that a condition on the angular size does not imply a selection on the physical size. The large proportion of rejected galaxies illustrates the challenge of carrying out confident morphology studies on large samples.

4.4. Correlations between the morphological indices

Of course, the different morphological indices are not statistically independent, sometimes even strongly correlated. This can in principle be explained by the fact that all of them are thought to accompany the Hubble sequence. The Pearson correlation coefficient measures the linear dependence between two variables. For more than two variables, the coefficients corresponding to all possible pairs of variables make up the correlation matrix, which is the covariance matrix for the centred and normalised variables.

The correlation coefficients between the morphological indices and the Hubble index are shown in Table 1. When the logarithm does not appear explicitly in the definition of index, which is the case for A , S , G and α , we take it in logarithmic scale. That corresponds to the behaviour in the plots and considerably increases some correlation coefficients. The bold numbers are the different combinations used as interaction indicators shown in Fig. 9. They are neither the most correlated nor the most uncorrelated combinations. The choice of two indices as interaction indicator is based on physical considerations. Galaxies that do not follow well-established correlations, or at the extremity of a weak correlation, are considered interacting. That explains why the coefficients of the chosen pairs are of intermediate values.

Such correlation measurements must be interpreted with care. The covariance is not a robust estimator of the dispersion of a sample, because it depends much on the outliers. Some indices show narrower relation than the correlation coefficient reveals. For instance, C and G are much more related (Lotz et al. 2004) than the correlation coefficient (0.62) might suggest. Other strong relations between morphological indices are already known in the literature, for example A - S (Conselice 2003; Conselice et al. 2003), or C - M_{20} (Lotz et al. 2004). Zamojski et al. (2007) study the relations between the five common indices and point out that a plane constituted by a concentration-like parameter (C or G) and a texture parameter (A or S) is the most convenient to bring out the colour bimodality. These two groups are highlighted by the correlation coefficients. C and G correlates with each other and are the best candidates to recover the Hubble sequence. As γ correlates well with A and S , it is clearly as a texture parameter; M_{20} appears to be an intermediate measure.

A high correlation coefficient shows that same information is sometimes contained in different indices which means that, for an automated classification, using all the indices would be redundant. In order to reduce the number of free parameters, the Zurich Estimator of Structure Types – ZEST (Scarlata et al. 2007) performs a principal component analysis (PCA) and maintains only the three first components, interpreting them as physical properties.

The anellipticity index seems to be quite independent of the other indices. This is explained by the fact that most galaxies, both early and late types, have a regular and thus elliptical outer shape. As could already be seen in Fig. 4, a strong correlation exists between our granulation and the clumpiness indices. Although defined differently, these two indices measure primarily the flux emitted by small scale structures. γ also correlates well with A and M_{20} , two indices interpreted as indicators of interactions or mergers.

5. Interaction chance indicator

It is very challenging and so far impossible to develop a completely automated method to reliably identify galaxies undergo-

Table 1. Correlation coefficients for the morphological catalogue.

	T^*	C	$\log(A)$	$\log(S)$	G	M_{20}	$\log(\alpha)$	γ
T^*	1.00							
C	-0.59	1.00						
$\log(A)$	0.36	-0.61	1.00					
$\log(S)$	0.47	-0.61	0.81	1.00				
G	-0.67	0.62	-0.11	-0.24	1.00			
M_{20}	0.40	-0.77	0.76	0.77	-0.33	1.00		
$\log(\alpha)$	-0.01	-0.09	0.36	0.34	0.20	0.25	1.00	
γ	0.32	-0.38	0.74	0.84	-0.11	0.64	0.38	1.00

*Hubble index.

ing galactic interactions and mergers. Even recent research on merging galaxies favours a visual inspection (Zamojski et al. 2011). The main difficulty is to define general criteria to create a segmentation map. In consequence of this arbitrary choice, two neighbouring galaxies could be considered one or two objects. A second difficulty is to photometrically estimate the physical properties specifying an interaction, such as the gas dynamics or a star formation indicator. For instance, the irregularity of the general shape of a galaxy or a disorderly distribution of the SF regions are a signature of interaction. But measuring these quantities is very sensitive to the observational conditions, which complicates the choice of the different parameters during the computational processes. These limitations explain why quantities determined by automated processes to recognise interaction or merger should be considered probability indicators. In this vein, we define a quantity that we call interaction chance indicator (I.C.I.²) based on the morphological indices. We prefer not to use the word probability because of the strong constraints imposed by the mathematical definition, especially the restriction of the allowed values to the range from 0 to 1. Thus, the I.C.I. is an arbitrary quantity which translates the probability for a galaxy to be in interaction.

For a long time, the asymmetry index has been seen as a simple – and the unique – morphological indicator of interacting galaxies (Conselice et al. 2000). Conselice (2003) improved the method by combining the asymmetry with the clumpiness index. They argue that, if a galaxy is positioned more than 3σ below a certain linear relation between these two indices, it can be interpreted as a merger. They identify this relation as $C = 0.35 S$.

We can build an interaction indicator by projecting the data along the relation direction onto the perpendicular straight line, as shown on the left of Fig. 9. The dotted lines follow the main relation of the two variables. They represent constant values of the interaction chance indicator and can so be regarded as equiprobabilities of interaction. Perpendicularly to the dotted lines, the arrows point out the direction of increasing interaction chance.

Lotz et al. (2004) designate as mergers the galaxies that do not follow the $M_{20} - G$ correlation. They roughly define a straight line, parallel to the main relation, above which they assume galaxies to be mergers. Figure 9 shows the $M_{20} - G$ plane (middle panel). The linear dependence of the two variables is calculated by a PCA excluding the outliers. The first PC is the last dotted line on the right. Perpendicular to it, the direction of the second PC is given by the arrow. An interaction chance indicator can be found by projecting the points onto the arrow along the dotted line.

² The anagram ICI is sometimes used for the inverse concentration index (Blanton et al. 2001).

With our method, we aim to identify the interacting galaxies through the distortions caused by the tidal forces. The anellipticity index was created for the purpose of quantifying the deformation of the outer shape of the galaxy. Nonetheless, α alone is not a good indicator of interaction, because it is very sensitive to irregularities near the Petrosian isophote, as mentioned above.

To avoid misinterpreting these artefacts, we require not only the shape but also the inner area of the galaxy to be irregular. This second criterion is measured by the asymmetry index. To make sure our I.C.I. is large only when both α and A have large values, we take the product and not the sum of these two indices. Therefore, our interaction chance indicator is defined as:

$$\text{I.C.I.} = A \cdot \alpha \quad (14)$$

In the $A - \alpha$ plane, the I.C.I. is the projection in a log-log plot, as seen on the right of Fig. 9. The I.C.I. can be obtained as the projection onto the arrow along the dotted lines. This time, we do not define the interaction indicator as a deviation from a main correlation, but as the upper boundary of a relation. Here, the dotted lines are not parallel to the correlation, but are constant values of the I.C.I. resulting from the definition.

Figure 10 shows galaxies from the morphological catalogue labelled as interacting. The surrounding line is the Petrosian isophote, evidencing that the studied system is sometimes actually composed of two or more galaxies.

When an interaction is still ongoing, the galaxies are effectively distorted and our indicator is efficient. But if the interaction is in its final phase, typically the end of the coalescence, the newly formed galaxy returns into a compact state with double or multiple nuclei, as exemplified by the last images of Fig. 10. In this case, our indicator is less efficient in detecting interaction.

Figure 11 shows galaxies from the same catalogue, not labelled as interacting but with a high I.C.I. Even not identified as such by the authors, some galaxies are obviously interacting. In some other cases, two or more galaxies are located angularly close to each other, which can or cannot be a sign of interaction. Unfortunately, in certain cases, the high value of the interaction chance indicator is entirely due to a field star. That is why our indicator must be interpreted as a chance of interaction.

In Table 2, we give the values of our I.C.I., as well as of the morphological indices and the morphological type for the first few galaxies of the three catalogues. For the complete table, refer to the electronic version of this paper.

Merger or interaction driven star formation can be very strong, as it is the case, for instance, in the ultra-luminous infrared galaxies (ULIRGs). On the other hand, BCDs have a very intense star formation in relation to their small masses. As BCDs are mainly non-interacting, the star formation is intrinsically triggered. Massive galaxies have a greater probability to be involved in interaction. Working on a small sample of HII galaxies,

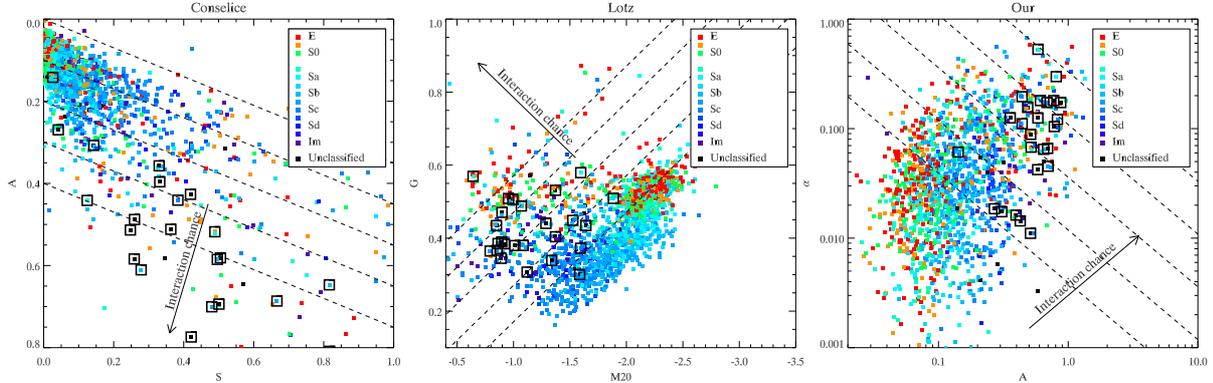


Fig. 9. Galaxies of the morphological catalogue in different planes used to determine interaction. Left: the $C - A$ plane used by Conselice (2003); Middle: the $M_{20} - G$ plane used by Lotz et al. (2004); Right: the $A - \alpha$ plane we use. The colour code indicates the morphological type as described by the legend. The overprinted black open squares are the galaxies identified as interacting by Fukugita et al. (2007). The I.C.I. is obtained by projecting the points onto the arrow along the dotted lines direction.

Table 2. Morphological indices, Interaction chance indicator (I.C.I.) and morphological type if available.

ID	C	A	S	G	M_{20}	α	γ	I.C.I.	T^*
Fukugita0001	3.407	0.089	0.016	0.520	-2.121	0.024	-1.978	0.00211	0.5
Fukugita0002	2.753	0.122	0.049	0.395	-2.117	-0.045	-1.723	-0.00550	3.0
Fukugita0003	3.181	0.095	0.003	0.507	-2.124	0.013	-2.024	0.00128	1.0
Fukugita0004	2.275	0.285	0.172	0.328	-1.458	0.019	-1.022	0.00544	3.5
Fukugita0005	2.259	0.061	0.026	0.230	-1.686	0.011	-1.994	0.00065	3.5
Fukugita0006	3.631	0.038	0.004	0.528	-2.218	0.003	-2.034	0.00011	0.0
...									
Kniazev0001	2.599	0.418	0.115	0.423	-1.901	0.029	-0.930	0.01203	—
Kniazev0002	2.170	0.358	0.751	0.341	-1.342	0.002	-1.585	0.00063	—
Kniazev0003	2.494	0.143	0.406	0.478	-1.712	0.005	-1.693	0.00076	—
...									
Izotov0001	2.690	0.770	0.353	0.549	-1.369	0.100	-0.588	0.07683	—
Izotov0002	2.529	0.594	0.808	0.464	-1.041	0.090	-0.510	0.05335	—
Izotov0003	2.767	0.123	0.026	0.402	-1.970	0.002	-1.557	0.00020	—
...									

*Hubble index by Fukugita et al. (2007).

This table is published in its entirety in the electronic edition.

Telles et al. (1997) shows that disturbed and extended galaxies have higher luminosities than compact and regular ones.

In HII galaxy catalogues, which are spectroscopically selected, we expect to encounter all kind of galaxies, small ones and large ones. Limiting our study to these objects, we should observe a relation between our I.C.I. and the physical size. The galactic diameter is given by the mean angular diameter of the Petrosian isophote and the galaxy's distance. The luminosity distance D_L is calculated from the redshift using a Λ CDM cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

Figure 12 shows the galactic diameter as a function of the interaction indicator for the Kniazev (blue squares) and Izotov (red diamonds) catalogues. The galaxies with a too low sharpness are drawn in light grey but are not considered in the calculation. The ellipse is the result of a PCA: the first PC follows the direction of the major axis (solid line) and the second PC is the minor axis (dotted line). The lengths of the axes of the ellipse are the 3σ dispersions of the PC. The first PC contains 57.7% of the total variance, i.e. a correlation coefficient of 0.16, what indicates a relation, but not a strong interdependence.

As the dwarf galaxies are compact and the massive galaxies subjected to more encounters, one could be surprised by such a

weak correlation between the size and the interaction indicator. In order to understand Figure 12, we chose two galaxies, indicated by asterisks, that deviate from the main correlation. Here, the small galaxy *Kniazev0368*, with a diameter of 1.65 kpc, has a slightly higher interaction indicator than the large *Kniazev0069*, which has a diameter of 38.25 kpc (Fig. 13).

Kniazev0069 has a regular and spherical low surface brightness halo, hardly perceptible in the image, which explains the large bump in the isophote on its right. This halo is rather blue, although the strong clumps are very blue. The galaxy is compact, but the irregular star formation distribution leads to quite large asymmetry and anellipticity indices, which implies a quite high I.C.I. value too.

Even if the weak resolution does not allow a clear diagnostic, it seems that *Kniazev0368* is in the final phase of a merger: it presents a double nucleus and low surface brightness relics of irregular structures. Both observations (Smith et al. 2007) and simulations (Di Matteo et al. 2007) show that mergers imply gas inflow into the central regions. At later stages of merger, the star formation takes place mainly in the centre, and the outer isophote is already less disturbed. In such conditions, our I.C.I.

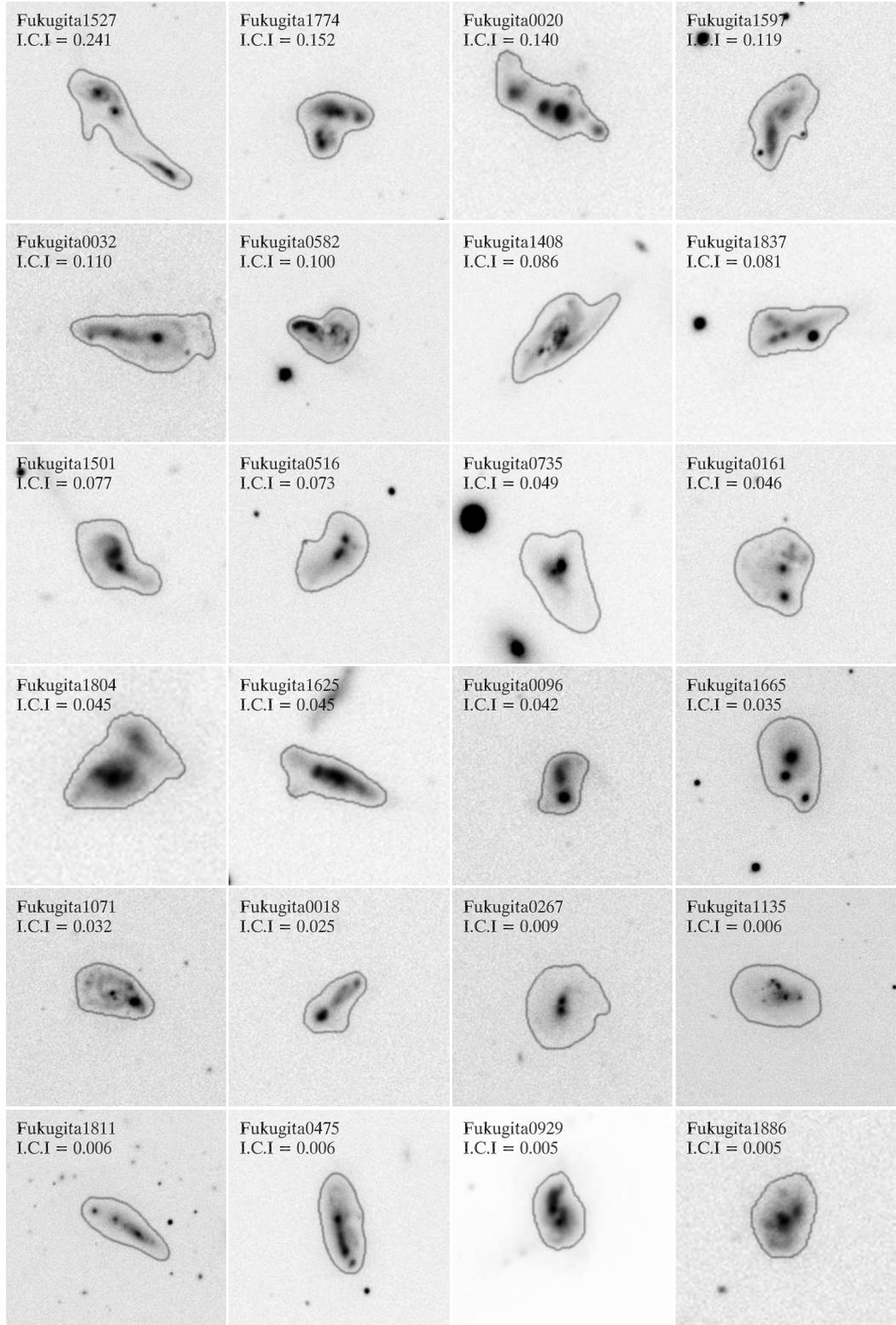


Fig. 10. Galaxies identified as interacting by Fukugita et al. (2007), drawn in order of decreasing interaction chance indicator. The surrounding line is the Petrosian isophote which determines the studied object. The images are in the g-band and in logarithmic grey-scale.

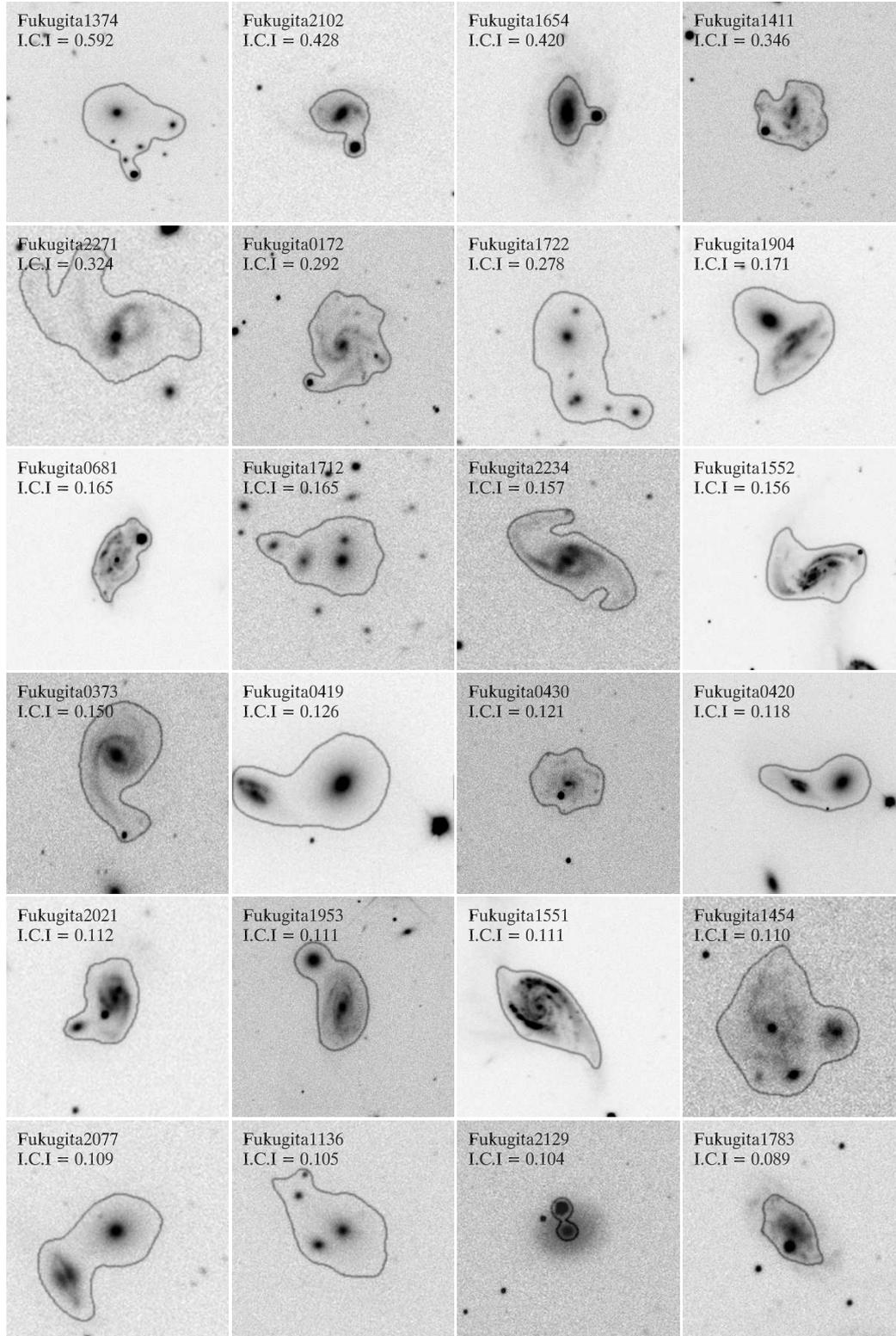


Fig. 11. Galaxies not identified as interaction by Fukugita et al. (2007), but presenting high interaction chance indicators. The images are in the *g*-band and in logarithmic grey-scale.

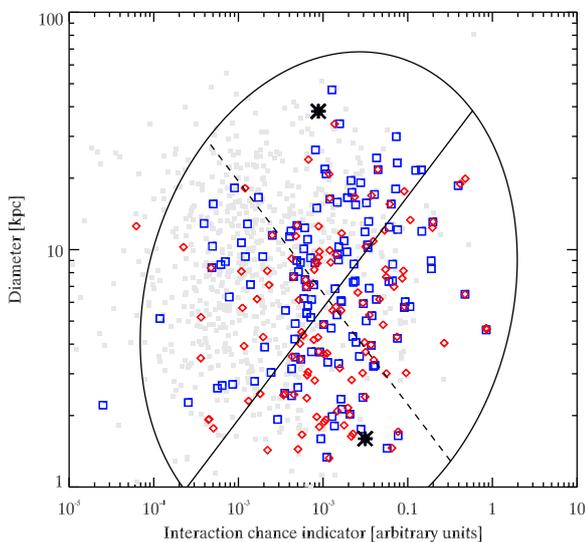


Fig. 12. Relation between our interaction chance indicator and the galactic diameter. The blue squares are the HII galaxies from the Kniazev catalogue and the red diamonds the ones from Izotov catalogue. The light grey points are galaxies with a sharpness less than 8, not well enough resolved for a reliable determination of the morphological indices. The ellipse has the two principal components with 3σ dispersions as axes. The two asterisks are the galaxies shown in Fig. 13.

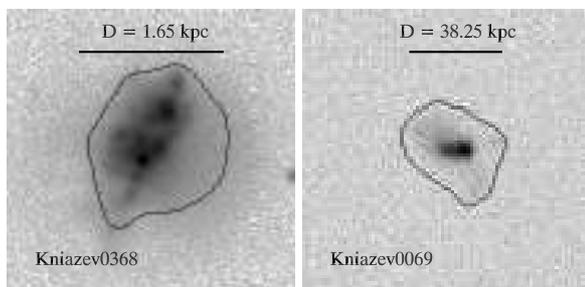


Fig. 13. *Kniazev0368* (left) has a higher interaction chance indicator than *Kniazev0069* (right).

will not have a high value. Actually, our indicator is sensitive to interaction and little to already ongoing mergers.

6. Discussion

We want to assess if merger driven starbursts are, on average, stronger than quiescent starbursts. The galaxies of the two HII catalogues were selected mainly according to their Balmer lines and they all have intense star formation. As the interacting galaxies are massive, they are expected to have higher SFR in absolute terms. That is why we prefer to compare the specific star formation rates (sSFR). The equivalent width of an ionised hydrogen line is a good estimator of the sSFR. In Fig. 14, we then plot the equivalent width of H β in function of the I.C.I. (left panel). Interacting galaxies have stronger star formation than non-interacting ones, but the tendency is weak: the correlation coefficient is only 0.14.

When the galaxies are resolved well enough, and that was the condition for our morphological study, one must be conscious of a recurrent bias in the joint photometric and spectroscopic studies. The fibre diameter of the SDSS spectrograph is 3 arcsec, that means 7.6 pixels of the images. For galaxies seen under a large angle, the spectra cover only a part of the galaxies. Frequently, the photometric pipeline of the SDSS chose the central part of the galaxy to get the spectrum. If one assumes the spectrum representative of the whole galaxy, the physical quantities derived from spectroscopic measures must be calibrated by the luminosities of the fibre and the galaxy. On the opposite, if one assumes the considered phenomenon occurring exclusively in the spectroscopic target, the raw measures can be used.

In our case, the spectroscopic fibre area is much smaller than the galaxy area, because we imposed a minimum resolution for confident morphological measurements. If the star formation takes place essentially in the centre, the spectrograph catches the entirety of the ionised-gas emission that the whole galaxy emits. But contrariwise, only a part of the stellar emission would be measured by the small aperture of the spectroscopic fibre, because it comes from the whole galaxy area and not only from the centre. The stellar emission forms the spectrum continuum; if it is undervalued, the equivalent widths of the emission lines, and consequently the sSFR, are overvalued. From this point of view, and despite of the high precision of the spectroscopic measurements, a photometric quantity is probably a better indicator of the total star formation.

To ensure not to suffer from such an aperture bias, we seek to express the sSFR from a photometric way. Although approximate, a blue band luminosity is an estimation of the recent and intermediate star formation (Gallagher et al. 1984) and the u -band can be considered as a SFR indicator (Moustakas et al. 2006; Hopkins et al. 2003). The red or infrared luminosity is dominated by the old population, that means proportional to the stellar mass. Thus, one can roughly estimate the sSFR from the colour $u-i$. On the right of Fig. 14, we plot $u-i$ in function of the I.C.I. for our two catalogues of HII galaxies. The relation between the I.C.I. and the sSFR also appears and the PCA shows an anticorrelation with the same correlation coefficient.

One could be surprised by such a weak correlation. Major mergers are known to be the most dramatic events during the life of a galaxy. The strongest star formation takes place during merger-driven starbursts.

A first explanation for this weak correlation could be that the strength of a starburst is not exactly measured by the sSFR, but rather by the star formation efficiency (SFE). The SFE is the ratio of the mass of the gas transformed into stars to the total gas mass, i.e. SFR/M_{gas} . The sSFR is proportional to the SFE only if the gas fraction is equal in all galaxies; if not, the SFE is inversely proportional to the gas fraction. In general, the gas fraction is far from constant and depends on the morphological type (Young & Scoville 1991), on the redshift (Tacconi et al. 2010) and on the stellar mass (Bell & de Jong 2000). In a local sample of HII galaxies, the dependences on the morphology and the redshift would probably not affect the gas fraction by much. But as less massive galaxies have higher gas fractions, we expect lower SFEs for BCD galaxies. Therefore, the SFE should show a better correlation with the I.C.I. than the sSFR does.

A second explanation would be that our I.C.I. is designed to assess the outline distortion, through the anellipticity index, and the non-symmetric flux distribution, through the asymmetry index. These two features are conspicuous in interacting galaxies, but not in remnants of mergers. During the early stages of a merger (first pass, maximum separation, second or third passes)

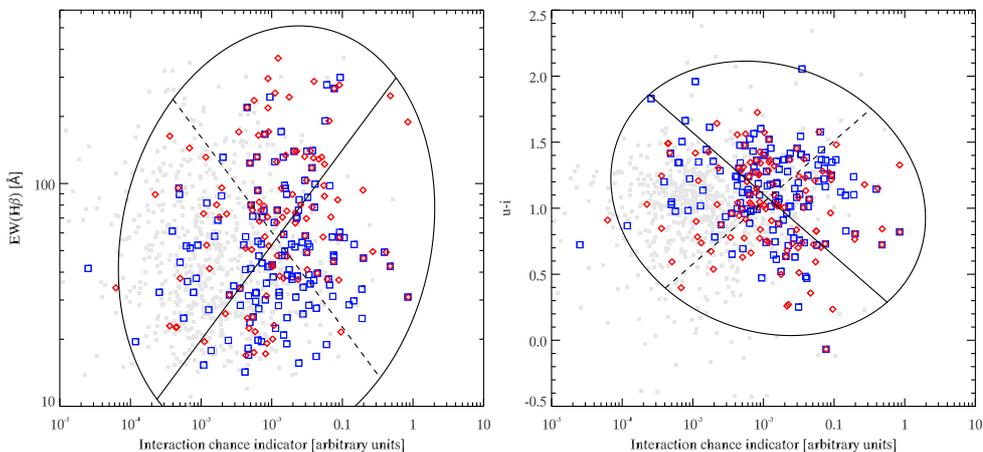


Fig. 14. Plot of the equivalent width of $H\beta$ and the colour $u-i$, both in function of our Interaction chance indicator. The coloured symbols and the ellipse are the same as in Fig. 12.

or galaxy flybys, the star formation is enhanced. Simulations teach us that, in a major merger, the main starburst is initiated only later, during the coalescence (Di Matteo et al. 2007; Cox et al. 2008). However, at late stages, the morphological signatures of interaction are already weaker. The most distant signature of dynamical disturbance, long tidal tails, quickly fade to low surface brightnesses. This way, our I.C.I. is more sensitive to the first stages of a merger event and no longer efficient to detect mergers in their last stages.

The granulation index γ takes on large values if the small scale structures, regarded as star forming regions, are bright and dispersed in the disc. We have just shown that the relationship between the I.C.I. and the star formation intensity is not very strong. Now, if the relation between the I.C.I. and γ would be narrower, this could be attributed to the geographical dispersion of the star forming regions. Figure 15 shows γ against the I.C.I. for the $H\text{II}$ galaxies of the Kniazev (blue squares) and Izotov (red diamonds) samples. The correlation coefficient is 0.39, much higher than in the previous figure.

We deduce that the star formation during the pre-merger can happen in the whole galaxy, not only in the central part. This does not prevent star formation, possibly very important, in the centre. In their simulation, Barnes & Hernquist (1996) already pointed out the importance of gas inflow in interacting galaxies, long before the final coalescence. And, observing interacting pairs in the infrared, Smith et al. (2007) detect gas migration already from the first gravitational disturbance.

The gas inflow becomes huge, especially in major mergers and, obviously, in mergers involving gas-rich disc galaxies (Di Matteo et al. 2007; Cox et al. 2008; Zamojski et al. 2011). Lotz et al. (2008, 2010b,a) simulated mergers and measured the common morphological indices on their model galaxies. They show that A takes on large values during the first and second passes, or at the beginning of the final merger. A similar behaviour occurs with G and M_{20} . Depending on the encounter configuration, the newly formed galaxy can turn quickly into a compact object, with an increasing C index. On the other hand, the SFR is highest in the final stage of the coalescence or even during the post-merger. It seems as if the star formation would respond to

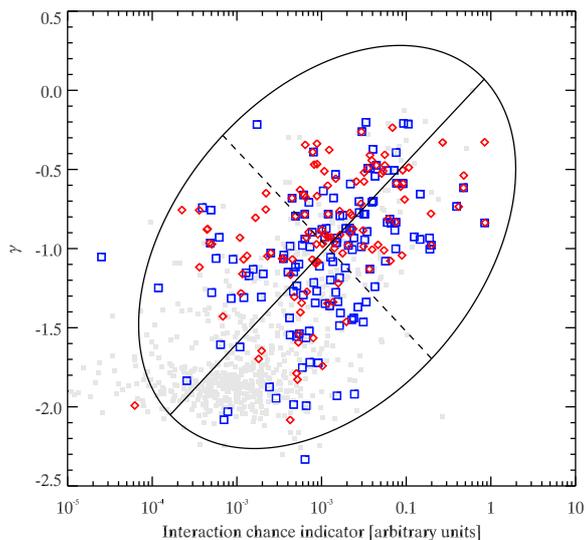


Fig. 15. Plot of the granulation index in function of our interaction chance indicator. The coloured symbols and the ellipse are the same as in Fig. 12.

the disturbed morphology with a delay. Consequently, there is no clear correlation between SFR and morphology during a merger.

In the pre-merger phase or during flybys, the interaction induces an enhanced SFR. The star formation at this stage takes place in the whole galaxy, in the centre and also in regions of the disc distant from the centre. On the other hand, during the coalescence phase, the star formation takes place exclusively in the centre. The SFR is maximal at this stage, the process needs a large fraction of the gas migrating to the centre, whereas the more peripheral gas, less dense and sheared, is no longer able to collapse.

The Schmidt-Kennicutt law (Schmidt 1959; Kennicutt 1989) relates the gas column density to the SFR column density. This law is no longer valid below a certain gas density threshold. The Toomre criterion (Toomre 1964) gives, for disc galaxies, a dynamical explanation for this limit: the rotational shear creates a velocity gradient, which results in a force that opposes the collapse. Thus, the presence of shearing turns a gas cloud more stable and suppresses the collapse unless the cloud is very dense.

While the merger is going on, dynamical disturbance creates an ever stronger shear in the outer structures. The Toomre criterion becomes more restrictive and the star formation outside the centre decreases with time. During this period, the star formation increases in the central region, being fuelled by gas inflow. At the last stage of the merger, the star formation takes place exclusively in the centre.

7. Summary

1. On a catalogue of visually classified galaxies, we computed the common morphological indices: concentration C , asymmetry A , clumpiness S , Gini coefficient G and second-order moment of the lightest pixels M_{20} . We defined the anellipticity index α as the distortion of the Petrosian isophote of the galaxy. We also defined the granulation index γ as the inclination of the wavelet transform spectrum weighted by the central second-order moment.
2. The study of the robustness of these indices inside each morphological type showed unconfident measurements for not well enough resolved images, when the angular diameter of the galaxy is less than 8 times the seeing. Some indices are affected by the signal-to-noise ratio or the ellipticity.
3. The correlation coefficients between the morphological indices and the Hubble type evince that the concentration-like parameters, C and G , are the most correlated to the Hubble sequence. A , S and γ are texture parameters, which means that they reveal the presence of clumps of light, physically interpreted as HII regions. M_{20} is somehow an intermediate parameter. As for α , it is neither a concentration nor a structure parameter because it was designed to consider only the shape of the border of the galaxy. It is, thereby, less related to the other indices.
4. Associated to A , α is used to define an interaction chance indicator (I.C.I) which is an arbitrary quantity reflecting the probability for a galaxy to be in interaction. Compared to similar approach using morphological indices, and by illustrating the Petrosian isophote on several examples of our morphological sample, our I.C.I. is more sensitive to the first stages of the interaction and less to the final merger phase.
5. Applying our method on two HII galaxy catalogues, we show that HII galaxies with a large I.C.I. have, on average, slightly larger diameter, slightly larger sSFR and notably larger granulation index γ .
6. We deduce that, at the beginning of an interaction, star forming regions appear in the whole galaxy and the morphological disturbance is maximum, but the starburst is not yet maximum. The SFR becomes maximum at a later stage, when dynamical instabilities are so important to cause massive gas inflow. At this stage, the star formation takes place essentially in the central region. This is consistent with recent simulations of galaxy encounters.

References

Abraham, R. G., van den Bergh, S., & Nair, P. 2003, *ApJ*, 588, 218

- Andrae, R., Jahnke, K., & Melchior, P. 2011a, *MNRAS*, 411, 385
 Andrae, R., Melchior, P., & Jahnke, K. 2011b, *ArXiv e-prints*
 Ball, N. M., Brunner, R. J., Myers, A. D., et al. 2008, *The Astrophysical Journal*, 683, 12
 Barnes, J. E. & Hernquist, L. 1996, *ApJ*, 471, 115
 Barnes, J. E. & Hernquist, L. E. 1991, *ApJ*, 370, L65
 Bastian, N. 2008, *MNRAS*, 390, 759
 Bell, E. F. & de Jong, R. S. 2000, *MNRAS*, 312, 497
 Bertin, E. & Arnouts, S. 1996, *A&AS*, 117, 393
 Blanton, M. R., Dalcanton, J., Eisenstein, D., et al. 2001, *AJ*, 121, 2358
 Conselice, C. J. 2003, *ApJS*, 147, 1
 Conselice, C. J., Bershad, M. A., Dickinson, M., & Papovich, C. 2003, *AJ*, 126, 1183
 Conselice, C. J., Bershad, M. A., & Jangren, A. 2000, *ApJ*, 529, 886
 Cox, T. J., Jonsson, P., Somerville, R. S., Primack, J. R., & Dekel, A. 2008, *MNRAS*, 384, 386
 Cuisinier, F., Moises, A. P., Pereira, M. L., Rabaça, C. R., & Gonçalves, D. R. 2005, in *American Institute of Physics Conference Series*, Vol. 804, *Planetary Nebulae as Astronomical Tools*, ed. R. Szczerba, G. Stasińska, & S. K. Gorny, 28–34
 Curty, D., Cuisinier, F. C., & Rabaça, C. R. 2010, in *IAU Symposium*, Vol. 262, *IAU Symposium*, ed. G. Bruzual & S. Charlot, 317–318
 Da Rocha, C. & Mendes de Oliveira, C. 2005, *MNRAS*, 364, 1069
 de Vaucouleurs, G. 1958, *ApJ*, 128, 465
 Di Matteo, P., Combes, F., Melchior, A.-L., & Semelin, B. 2007, *A&A*, 468, 61
 Doi, M., Fukugita, M., & Okamura, S. 1993, *MNRAS*, 264, 832
 Epitácio Pereira, D. N., de Mello, A. B., & Rabaça, C. R. 2007, in *IAU Symposium*, Vol. 235, *IAU Symposium*, ed. F. Combes & J. Palous, 233–233
 Fukugita, M., Nakamura, O., Okamura, S., et al. 2007, *AJ*, 134, 579
 Gallagher, III, J. S., Hunter, D. A., & Tutukov, A. V. 1984, *ApJ*, 284, 544
 Gini, C. 1912, *Variabilità e mutabilità*, ed. Gini, C. (Libreria Eredi Virgilio Veschi, Roma)
 Goddard, Q. E., Bastian, N., & Kennicutt, R. C. 2010, *MNRAS*, 405, 857
 Hopkins, A. M., Miller, C. J., Nichol, R. C., et al. 2003, *ApJ*, 599, 971
 Hunter, D. A. & Elmegreen, B. G. 2004, *AJ*, 128, 2170
 Izotov, Y. I., Stasińska, G., Meynet, G., Guseva, N. G., & Thuan, T. X. 2006, *A&A*, 448, 955
 Kelly, B. C. & McKay, T. A. 2004, *AJ*, 127, 625
 Kennicutt, Jr., R. C. 1989, *ApJ*, 344, 685
 Kent, S. M. 1985, *ApJS*, 59, 115
 Kniazev, A. Y., Pustilnik, S. A., Grebel, E. K., Lee, H., & Pramskij, A. G. 2004, *ApJS*, 153, 429
 Kushner, L. K., Obric, M., West, A. A., & Dalcanton, J. 2006, in *Bulletin of the American Astronomical Society*, Vol. 38, *American Astronomical Society Meeting Abstracts*, 1037–+
 Lauer, T. R. 1985, *MNRAS*, 216, 429
 Leal Ferreira, M. L., Rabaça, C. R., Cuisinier, F., & Epitácio Pereira, D. N. 2009, in *Revista Mexicana de Astronomía y Astrofísica Conference Series*, Vol. 35, *Revista Mexicana de Astronomía y Astrofísica Conference Series*, 290–+
 Lisker, T. 2008, *ApJS*, 179, 319
 Lisker, T., Debattista, V. P., Ferreras, I., & Erwin, P. 2006, *MNRAS*, 370, 477
 Lorenz, M. O. 1905, *Publications of the American Statistical Association*, Volume 9, Number 70, p. 209–219, 9, 209
 Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2008, *MNRAS*, 391, 1137
 Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2010a, *MNRAS*, 404, 590
 Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2010b, *MNRAS*, 404, 575
 Lotz, J. M., Primack, J., & Madau, P. 2004, *AJ*, 128, 163
 Lupton, R., Gunn, J. E., Ivezić, Z., Knapp, G. R., & Kent, S. 2001, in *Astronomical Society of the Pacific Conference Series*, Vol. 238, *Astronomical Data Analysis Software and Systems X*, ed. F. R. Harnden Jr., F. A. Primini, & H. E. Payne, 269–+
 Massey, R. & Refregier, A. 2005, *MNRAS*, 363, 197
 Melchior, P., Meneghetti, M., & Bartelmann, M. 2007, *A&A*, 463, 1215
 Moustakas, J., Kennicutt, Jr., R. C., & Tremonti, C. A. 2006, *ApJ*, 642, 775
 Naim, A., Ratnatunga, K. U., & Griffiths, R. E. 1997, *ApJ*, 476, 510
 Ngan, W., van Waerbeke, L., Mahdavi, A., Heymans, C., & Hoekstra, H. 2009, *MNRAS*, 396, 1211
 Odewahn, S. C., Cohen, S. H., Windhorst, R. A., & Philip, N. S. 2002, *ApJ*, 568, 539
 Parodi, B. R. & Binggeli, B. 2003, *A&A*, 398, 501
 Pereira, D. N. E. & Rabaça, C. R. 2003, *Bulletin of the Astronomical Society of Brazil*, 23, 33
 Petrosian, V. 1976, *ApJ*, 209, L1
 Refregier, A. 2003, *MNRAS*, 338, 35
 Scarlata, C., Carollo, C. M., Lilly, S., et al. 2007, *ApJS*, 172, 406
 Schade, D., Lilly, S. J., Crampton, D., et al. 1995, *ApJ*, 451, L1+
 Schmidt, M. 1959, *ApJ*, 129, 243
 Seiden, P. E. & Schulman, L. S. 1990, *Advances in Physics*, 39, 1

- Sérsic, J. L. 1963, Boletín de la Asociación Argentina de Astronomía La Plata Argentina, 6, 41
- Shensa, M. J. 1992, IEEE Transactions on Signal Processing, 40, 2464
- Smith, B. J., Struck, C., Hancock, M., et al. 2007, AJ, 133, 791
- Starck, J.-L., Murtagh, F. D., & Bijaoui, A. 1998, Image Processing and Data Analysis, ed. Starck, J.-L., Murtagh, F. D., & Bijaoui, A. (Cambridge University Press)
- Tacconi, L. J., Genzel, R., Neri, R., et al. 2010, Nature, 463, 781
- Takamiya, M. 1999, ApJS, 122, 109
- Telles, E., Melnick, J., & Terlevich, R. 1997, MNRAS, 288, 78
- Toomre, A. 1964, ApJ, 139, 1217
- Toomre, A. & Toomre, J. 1972, ApJ, 178, 623
- van der Wel, A. 2008, ApJ, 675, L13
- Yamauchi, C., i. Ichikawa, S., Doi, M., et al. 2005, AJ, 130, 1545
- York, D. G., Adelman, J., Anderson, J. J. E., et al. 2000, AJ, 120, 1579
- Young, J. S. & Scoville, N. Z. 1991, ARA&A, 29, 581
- Zamojski, M., Yan, L., Dasyra, K., et al. 2011, ApJ, 730, 125
- Zamojski, M. A., Schiminovich, D., Rich, R. M., et al. 2007, ApJS, 172, 468

Appendix A: The wavelet transform and the *à trous* algorithm

Starck et al. (1998) give a very clear and extended mathematical description of the wavelet transform (WT), as well as its computational implementation. Please, refer to their work for a complete introduction. We just present here a short overview of the method.

A wavelet is a localised function of average value zero and of compact support. There are many kinds of wavelets, but they are all wave-like clipped to a finite domain. The WT consists of projections of the input data onto a set of wavelets whose spatial frequency varies. A family of wavelets $w_a(x)$ is generated by dilatation of the mother wavelet $w(x)$ by a scale factor $a > 0$:

$$w_a(x) = \frac{1}{\sqrt{a}} w\left(\frac{x}{a}\right) \quad (\text{A.1})$$

The wavelets $w_a(x)$ act as pass-band filters for scales between a and $2a$. A one-dimensional function $f(x)$ can be expressed as a linear combination of wavelets. The continue WT is defined by the wavelet coefficients $c_a(k)$ which are the convolutions of $f(x)$ by the wavelets $w_a(x)$:

$$c_a(k) = \int_{-\infty}^{\infty} f(x) w_a(x - k) dx \quad (\text{A.2})$$

$$\stackrel{\text{def}}{=} (f * w_a)(k) \quad (\text{A.3})$$

where k is the position parameter and $*$ indicates the convolution product.

The coefficients $c_a(k)$ measure the fluctuations of the signal $f(x)$ around the point k and at the scale a . This shows that the WT, unlike the Fourier transform, does not lose the geographical informations and permits to extract information about both the positions and the dimensions of the contained sub-structures.

In a dyadic approach, only scale factors as powers of 2 are considered. We set $a = 2^i$; i is sometimes called the ‘‘octave’’. For a CCD image, i.e. a two-dimensional digital signal, the functions must be discretised and the wavelet coefficients are obtained by a convolution with a bi-dimensional wavelet as kernel. Mathematically, the wavelet coefficients of the image $f[x, y]$ are then defined as following:

$$c_i[k, l] = (f * w_i)[k, l] \quad (\text{A.4})$$

$$= \frac{1}{2^i} \sum_{x=0}^{N_1} \sum_{y=0}^{N_2} f[x, y] w\left[\frac{x}{2^i} - k, \frac{y}{2^i} - l\right]. \quad (\text{A.5})$$

In practice, to spare many convolutions – very expensive in term of computational time –, we calculate the coefficients in a different way. The filter bank of the wavelets can more easily be created by using high-pass filters. This means, instead of choosing a wavelet function, we chose a scaling function $v(x)$ and the wavelet corresponding to the scaling function will *a posteriori* be associated to it. The choice of a scaling function $v(x)$ completely defines the wavelet $w(x)$. From the scaling relation A.1, one can take advantage of the recursive property and build the wavelet coefficients by subtracting two successive smoothed images.

$$c_i[k, l] = (f * v_i)[k, l] - (f * v_{i+1})[k, l] \quad (\text{A.6})$$

The *à trous*³ algorithm (Shensa 1992; Starck et al. 1998) leads to a redundant WT, as no decimation is carried out. At each step, the scaling function is dilate by inserting zeros, namely holes. The result is, that the wavelet coefficients $c_i(k, l)$ have the same size as the original image, which is obviously redundant, but greatly facilitates the multi-scale analysis on the other hand.

To ease and speed up the computation, we assume for the convolution the separability of the two dimensions. As scaling function, we choose the B_3 -spline. Thus, at the first step, we perform a row-by-row followed by a column-by-column convolution with the kernel

$$v_0 = \left[\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16} \right]. \quad (\text{A.7})$$

In the second step, we perform onto the previous result a new convolution with the kernel

$$v_1 = \left[\frac{1}{16}, 0, \frac{1}{4}, 0, \frac{3}{8}, 0, \frac{1}{4}, 0, \frac{1}{16} \right]. \quad (\text{A.8})$$

And so on, until the scale reaches the size of the image. The boundaries are handled by assuming continuity of the sky: the last pixel is repeated. The associated wavelet to the *à trous* algorithm with the B_3 -spline as scaling function can not be written analytically. However, when n tends to infinity, the cardinal B-spline B_n tends to the Gaussian function. That means that, for large octaves, the wavelet tends to a difference of Gaussians, a narrow one subtracted by a twice wider one. Note that it is close but dissimilar from the Mexican hat. Figure A.1 shows the X-Z-plane of the first four wavelets. As indicated by the different ordinate axis scales, the amplitude of the wavelet decreases with increasing octave.

In the WT coefficients, the level of relevance is chosen at 5σ of the sky noise. This threshold, different for each scale, defines the zone of interest in the wavelet coefficients; outside these areas, the pixels are set to zero. Figure A.2 shows the wavelet transform of the galaxy *Fukugita2244*. The lines are the isophotes defining the edges of the zones of interest.

The WT segregates the structures in function of their size. The first wavelet coefficient contains only elements at a scale of 1 pixel, the second coefficient at a scale of 2 pixels, the third at 4 pixels and so on. The structures highlighted in the coefficients are not physical objects, but only regions with more light than the neighbourhood. At very small scales, they are mostly HII-regions; at larger scales, they can be spiral arms, the bulge, or the entire disk.

³ ‘‘à trous’’ is the french expression for ‘‘with holes’’.

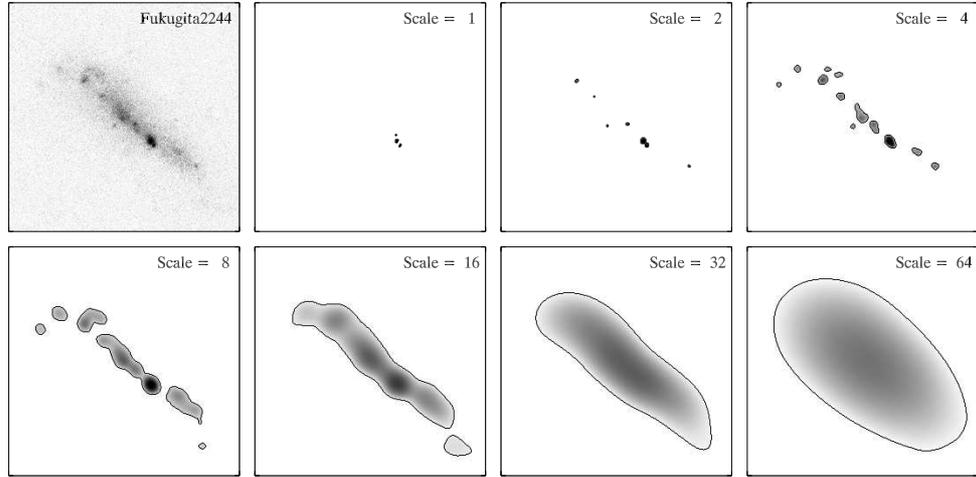


Fig. A.2. Wavelet transform of the galaxy *Fukugita2244*. The first panel is the original image, and the other panels are the wavelet coefficients. The lines are the isophotes at 5σ of the noise and define the zones of interest. Images are shown in the same logarithmic scale.

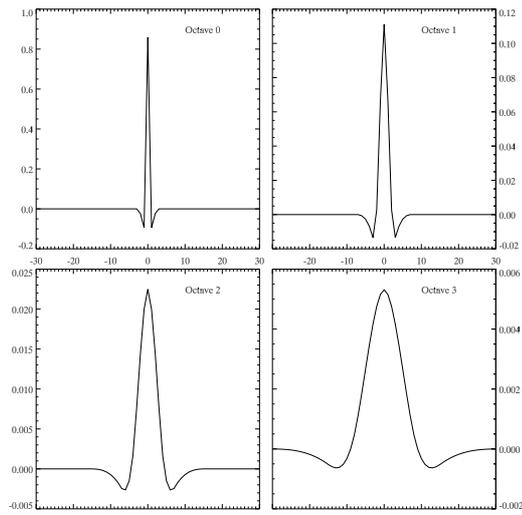


Fig. A.1. The first four wavelets associated to the B_3 -spline as scaling function used with the *à trous* algorithm. The unit of the X-axis is a pixel.

5. ANELLIPTICITY AND GRANULATION AS MORPHOLOGICAL INDICES TO EXPLORE STAR FORMATION IN INTERACTING GALAXIES.

Capítulo 6

Gas and stellar metallicities in HII galaxies

Este artigo, publicado no MNRAS (*Monthly Notices of the Royal Astronomical Society*), estuda a evolução da metalicidade das galáxias HII. Todos os espectros galácticos do SDSS foram considerados, mas somente aqueles com uma razão sinal/ruído suficiente alta, com linhas de emissão fortes e com a presença da linha [OII] $\lambda 3727 \text{ \AA}$ foram conservados. Essa última condição exclui as galáxias com um *redshift* menor do que $z_{min} = 0.033$. De um total de $\sim 930\,000$ espectros do SDSS, apenas 712 entraram em nossa amostra final de galáxias HII. Uma outra amostra, menos restritiva e chamada de amostra teste, contém dez vezes mais galáxias com linhas de emissão.

Depois de haver comparado onze métodos diferentes de determinação das abundâncias por linhas de emissão, a metalicidade do gás é avaliada seguindo Pilyugin & Thuan (2005). Esta metalicidade é fixada para a população estelar jovem. Pelo método de síntese de população, modelamos as galáxias como sendo compostas por uma população velha, uma intermediária e uma jovem, para determinar as massas, idades e metalicidades das populações parciais.

Os nossos resultados mostram um comportamento diferente entre a amostra teste e a amostra de galáxias HII.

A amostra teste segue a relação massa-metalicidade, ou seja, a metalicidade do gás aumenta com a massa da galáxia. Isso indica que as galáxias de baixa massa são quimicamente menos evoluídas do que as galáxias de alta massa. Ao contrário, o enriquecimento da metalicidade, definido como a diferença entre a metalicidade da

6. GAS AND STELLAR METALLICITIES IN HII GALAXIES

população jovem e a da população velha, não mostra dependência sistemática com a massa galáctica. O que traduz o fato de que, em média, o mecanismo de reciclagem do meio interestelar é o mesmo nas galáxias de alta e baixa massas.

Na amostra de galáxias HII, por outro lado, não há dependência entre a metalicidade do gás e a massa. Mas para o enriquecimento da metalicidade, reparamos a existência de uma fraca dependência para massas maiores do que $5 \cdot 10^9 M_{\odot}$. Interpretamos esse comportamento com ajuda de resultados recentes de simulações hidrodinâmicas de galáxias anãs irregulares. Nessas simulações, dois mecanismos reduzem o enriquecimento químico: (1) A expulsão do gás rico em metal, recentemente formado na última geração estelar pelos ventos galácticos (*gas outflow*) e (2) a queda de gás do meio intra-aglomerado pobre em metal (*gas infall*). Embora cada um dos dois processos tenha como efeito diminuir a metalicidade da galáxia, a combinação dos dois tem uma eficiência menor. De fato, o gás pobre que cai impede a expulsão do gás rico. Portanto, as nossas observações favorecem modelos com constante queda de gás. Nas galáxias com massas maiores do que $5 \cdot 10^9 M_{\odot}$, o potencial gravitacional é suficiente para reter o gás rico em metal e, conseqüentemente, a metalicidade aumenta mais rapidamente.

Gas and stellar metallicities in H II galaxies

P. Westera,^{1,2*} F. Cuisinier,^{3†} D. Curty^{2*} and R. Buser^{4*}

¹Universidade Federal do ABC, Rua Santa Adélia 166, 09.210-170 Santo André, SP, Brazil

²Observatório Nacional, Rua José Cristino 77, 20.921-400 Rio de Janeiro, RJ, Brazil

³Observatório do Valongo/UFRJ, Ladeira do Pedro Antônio 43, 20.080-090 Rio de Janeiro, RJ, Brazil

⁴Departement Physik, Universität Basel, Klingelbergstrasse 82, 4056 Basel, Switzerland

Accepted 2011 December 1. Received 2011 November 26; in original form 2011 March 29

ABSTRACT

We examine the gas and stellar metallicities in a sample of H II galaxies from the Sloan Digital Sky Survey, which possibly contains the largest homogeneous sample of H II galaxy spectra to date.

We eliminated all spectra with an insufficient signal-to-noise ratio, without strong emission lines and without the [O II] $\lambda 3727$ Å line, which is necessary for the determination of the gas metallicity. This excludes galaxies with redshift $\lesssim 0.033$. Our final sample contains ~ 700 spectra of H II galaxies.

Through emission line strength calibrations and a detailed stellar population analysis employing evolutionary stellar synthesis methods, which we already used in previous works, we determined the metallicities of both the gas and the stellar content of these galaxies.

We find that in H II galaxies up to stellar masses of $5 \times 10^9 M_{\odot}$, enrichment mechanisms do not vary with galactic mass, being the same for low- and high-mass galaxies on average. They do seem to present a greater variety at the high-mass end, though, indicating a more complex assembly history for high-mass galaxies. In around 23 per cent of our H II galaxies, we find a metallicity decrease over the last few Gyr. Our results favour galaxy evolution models featuring constantly infalling low-metallicity clouds that retain part of the galactic winds. Above $5 \times 10^9 M_{\odot}$ stellar mass, the retention of high-metallicity gas by the galaxies' gravitational potential dominates.

Key words: ISM: abundances – galaxies: abundances – galaxies: evolution – galaxies: starburst.

1 INTRODUCTION

H II galaxies are characterized by prominent emission lines. In fact, they are defined as having strong H β and O III lines, but other ionized hydrogen (H α , H γ , etc.) and high excitation metallicity lines (O II, N II and others) are strong too. These emission lines are driven by photoionization of interstellar gas by hot massive stars from young stellar populations (Sargent & Searle 1970; French 1980). The star-forming activity in H II galaxies is so strong that it can certainly not have been maintained at its present level during a Hubble time (see e.g. Searle, Sargent & Bagnuolo 1973). In fact, they are the most extreme case of star-forming galaxies, showing the highest excitation emission lines, and thus the highest (relative) star formation rates (SFRs).

They are also the lowest metallicity galaxies of the interstellar medium, with metallicities around one-tenth solar (Pérez-Montero

& Díaz 2003; Vílchez & Iglesias-Páramo 2003; Kniazev et al. 2004; Izotov et al. 2006; Kehrig et al. 2006; Hägele et al. 2008; Pérez-Montero et al. 2010, and others).

It is now clear that the hypothesis of H II galaxies as galaxies experiencing their first star formation burst can be ruled out, and that their stellar content is predominantly old, older than 1 Gyr (Raimann et al. 2000; Cid Fernandes, Leão & Lacerda 2003; Kong et al. 2003; Westera et al. 2004; Asari et al. 2007; Cid Fernandes et al. 2007; Hoyos et al. 2007, and others). Nevertheless, both the high relative SFRs and the low metallicities indicate that H II galaxies are among the least evolved galaxies in existence, which assigns to them a special role as fossil record of galaxy evolution (Lequeux et al. 1979). They are, therefore, ideal objects for chemical evolution studies.

Several authors have developed theoretical evolutionary models of dwarf irregular galaxies, of which H II galaxies are a subcategory. Some of these models predict the galaxies' present-day stellar and gas metallicities, among other properties. In these models, several chemical enrichment processes are identified.

MacLow & Ferrara (1999) developed single-phased hydrodynamical models for dwarf galaxies. They model the effects of

*E-mail: pieter.westera@ufabc.edu.br, pieter@on.br (PW); curty@on.br (DC); Roland.Buser@unibas.ch (RB)

†In memoriam (1969–2011).

repeated Type II supernova (SNII) explosions from starbursts on the interstellar medium, i.e. the enrichment by the ejected metals on the one hand, and the gas loss by supernova winds on the other hand, taking into account the gravitational potential of their dark matter haloes. They find that, in galaxies with gas masses below $10^6 M_{\odot}$, most of the gas is blown away, and in galaxies with masses between 10^7 and $10^9 M_{\odot}$, mainly the newly formed metal-rich gas is blown away, whereas the already present, metal-poor, gas is retained. They do not calculate the final (present-day) metallicities of their model galaxies, but it is obvious that they remain low metallicity, at least the ones with gas masses below $10^9 M_{\odot}$.

Recchi, Matteucci & D’Ercole (2001) and Recchi et al. (2002) simulate IZw18, one of the lowest metallicity galaxies known, using a single star formation burst model and a model with a doubly peaked star formation history. They use a lower SNII heating efficiency than MacLow et al. in their models, but include Type Ia supernovae (SNIa), for which they use a higher heating efficiency. As a consequence, alpha elements are ejected less efficiently than in the MacLow et al. model, resulting in higher metallicities. The iron-peak elements, which are produced in SNIa, on the other hand, are still ejected, causing high $[\alpha/\text{Fe}]$ ratios. The (gas) metallicity in their single burst model remains very low, $[\text{O}/\text{H}] \lesssim -1.3$, depending on the galaxy mass. In their (more likely) double burst model, $[\text{O}/\text{H}]$ can reach values up to -1 , depending on the details of the models, i.e. the stellar yields used, the stellar initial mass function (IMF), the time between the bursts, the duration of the second burst, the gas density and metallicity in the star-forming region, and the total mass of the stars formed in the second burst. The final metallicities of the double burst model galaxies are between 0.6 and 1 dex higher than after the first burst.

Tenorio-Tagle et al. (2006) include photoionization and cluster wind in their two-dimensional hydrodynamic calculations. They identify two mechanisms: the storage of clouds into a long-lasting ragged shell inhibiting the expansion of the thermalized wind and the steady filtering of the shocked wind gas through channels carved within the cloud layer. They conclude that both processes must be at work in H II galaxies. Unfortunately, they make no prediction about the present-day metallicity of their model galaxies.

In a series of papers, Recchi and collaborators use a new generation of chemo-dynamical models, some of which include infalling clouds. The first two, Recchi et al. (2004, 2006), are aimed at reproducing the properties of two specific objects, IZw18 and NGC1569, respectively. The third one, Recchi & Hensler (2007) investigates the influence of several factors, such as the cloudiness of the gas distribution, the IMF slope and stellar yields. They find that models with continuous low-level star formation periods in the past followed by a quiescent phase and a recent, stronger burst best reproduce the chemical properties of the studied galaxies. They also find that, for a homogeneous gas distribution, metals get blown away by galactic winds, but cool infalling intergalactic clouds can hamper these galactic winds. The final metallicities in their various models vary from -2.5 to -0.8 .

Although metallicity determinations of the various components of H II galaxies cannot discriminate how much each of these mechanisms contribute to the chemical evolution, they can quantify the metallicity at various stages of a galaxy’s evolution, i.e. at the time of the formation of the old (> 1 Gyr) stellar populations and at present, and thereby help to rule out some of the evolutionary scenarios, and support others. H II galaxies have the advantage that it is possible to determine independently the metallicities of the young and the old populations from their spectra. As the young populations were formed recently out of the same gas that is responsible for the emis-

sion lines, their metallicity can be assumed to be more or less the same as the one of the gas, which can be determined from the emission line strengths. The metallicity of the old populations, on the other hand, can be determined from the continuum and absorption features through population synthesis.

The main goal of the present work is to measure the chemical enrichment in a homogeneous sample of H II galaxies, and examine possible trends with galactic properties such as total mass. This way, we can evaluate the different models and make statements on the importance of the various enrichment mechanisms.

Until recently, the discovery of H II galaxies was limited to the visual inspection of objective-prism surveys, introducing ill-controlled biases and selection effects (for compilations, see Terlevich et al. 1991; Kehrig, Telles & Cuisinier 2004, and references therein).

The Sloan Digital Sky Survey (SDSS; York et al. 2000) presents for the first time a comprehensive data base of galactic spectra. Data release 7 (Abazajian et al. 2009) contains over 900 000 galaxy spectra that have been selected on clear magnitude limiting criteria. In an unprecedented manner, the SDSS allows us to study star-forming emission line galaxies based on clear and quantitative criteria, and not mere visual inspection.

In this work, we define clear emission line strength criteria to distinguish galaxies containing strongly excited gas from other emission line galaxies. We only include spectra that contain the $[\text{O II}] \lambda 3727 \text{ \AA}$ emission line, as this line is indispensable for an adequate determination of the gas metallicity. This criterion excludes galaxies with redshift $\lesssim 0.033$. This way, we selected a homogeneous sample of ~ 700 H II galaxy spectra of high quality from the SDSS.

In this sample, we determine the metallicities of both the gas and the old stellar content using emission line calibration and stellar population synthesis methods, respectively. It is the first time an independent metallicity determination of the gas and the stars is performed on a clearly defined H II galaxy sample.

The layout of this article is the following. In Section 2, the catalogue of spectra analysed in this work is presented. Section 3 gives a detailed description of the method we used to analyse the spectra. Subsection 3.1 describes how we determined the gas metallicity, and Section 3.2 is dedicated to the determination of the stellar metallicities. The main results are given in Section 4 and discussed in Section 5. A summary and the main conclusions can be found in Section 6.

2 THE DATA BASE

For this work, we used spectra from the SDSS Data Release 7, which have a signal-to-noise ratio (S/N) of $S/N > 4 \text{ pixel}^{-1}$ at $m_g = 20.2$, but usually much higher, and cover a wavelength range 3800–9200 Å. Most important of all, it contains a large number of galaxy spectra, that is around 900 000.

Any analysis however relies on previous clear criteria to define what H II galaxies are. First, we rejected all spectra whose wavelength ranges do not include the region from 3677 to 3775 Å, which is needed to measure the $[\text{O II}] \lambda 3727 \text{ \AA}$ emission line strength, used in many strong line methods to determine the gas metallicity. This excludes galaxies with redshifts below 0.033. Another prerequisite for measuring this line strength is a sufficient S/N in this wavelength region. We only kept spectra with $S/N_{[\text{O II}]\lambda 3727} \geq 2.8$. This criterion guarantees sufficiently low noise for the entire spectral range used in this work, since at longer wavelengths, the S/N of the spectra is higher than around the $[\text{O II}] \lambda 3727 \text{ \AA}$ line, i.e. $\gtrsim 8$ in g.

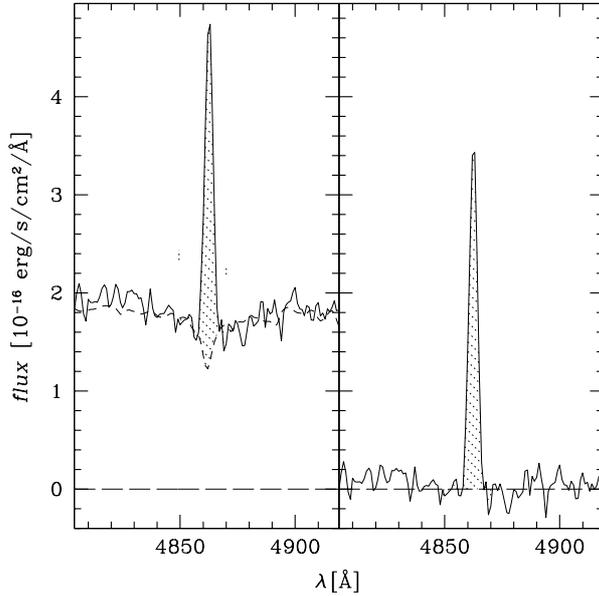


Figure 1. Illustration of the emission line strength measurement (here $H\beta$ in 51957–0273–418). In the left-hand panel, the solid line represents the observed spectrum (rebinned to the BC03 wavelength grid), whereas the short-dashed line shows the best fit using the BC03 high-resolution library. The shaded region between these two lines shows the area used to calculate the emission line strength. The right-hand panel shows the emission line after subtracting the best fit. The shaded area corresponds to the shaded area between spectrum and best fit from the left-hand panel.

The spectra were then corrected for Galactic (foreground) gas extinction using the values given in the SDSS data base, which were derived from the Schlegel, Finkbeiner & Davis (1998) reddening maps. Using these extinction values and the Galactic extinction law of Fitzpatrick (1999), we dereddened the spectra. After that, we deredshifted them, and measured the emission line strengths.

In order to properly determine the emission line strengths, we first had to remove the contribution from the absorption lines of the underlying stellar populations. This was done by subtracting high-resolution spectra representing the stellar continuum from the empirical spectra as illustrated in Fig. 1. These high-resolution spectra were compiled by fitting the spectra of composite stellar populations, made up of three single stellar populations (SSPs), a young (≤ 10 Myr), an intermediate age (20–500 Myr) and an old one (5 Gyr), from the ‘BC03’ integrated spectral energy distribution (ISED) library to the continua of the galaxy spectra. The ‘BC03’ library was produced using the Bruzual and Charlot (2003) Galaxy Isochrone Spectral Synthesis Evolution Library (GISSEL) code (Charlot & Bruzual 1991; Bruzual & Charlot 1993, 2003) implementing the Padova 1995 isochrones (Fagotto et al. 1994; Girardi et al. 1996) combined with the STELIB (Le Borgne et al. 2003) stellar library. Since it is made up of empirical high-resolution spectra, the ‘BC03’ library reproduces absorption line shapes well, and is ideal for this purpose. Table 1 lists all the lines we measured.

However, incomplete coverage of the stellar parameter space and calibration uncertainties of the STELIB library result in systematic errors in the overall spectral shapes. Therefore, the ‘BC03’ library is not a good choice for determining the properties of the stellar content from a full spectral fit. For the stellar population synthesis,

Table 1. Emission lines measured in this work.

Line	Central wavelength
[O II] $\lambda 3727 \text{ \AA}$	3727 \AA
H ϵ	3970 \AA
H δ	4102 \AA
H γ	4340 \AA
[O III] $\lambda 4363 \text{ \AA}$	4363 \AA
[He II] $\lambda 4686 \text{ \AA}$	4686 \AA
H β	4861 \AA
[O III] $\lambda 4959 \text{ \AA}$	4959 \AA
[O III] $\lambda 5007 \text{ \AA}$	5007 \AA
[He I] $\lambda 5876 \text{ \AA}$	5876 \AA
[O I] $\lambda 6300 \text{ \AA}$	6300 \AA
[S III] $\lambda 6312 \text{ \AA}$	6312 \AA
[N II] $\lambda 6548 \text{ \AA}$	6548 \AA
H α	6563 \AA
[N II] $\lambda 6584 \text{ \AA}$	6584 \AA
[S II] $\lambda 6717 \text{ \AA}$	6717 \AA
[S II] $\lambda 6731 \text{ \AA}$	6731 \AA
[O II] $\lambda 7319 \text{ \AA}$	7319 \AA
[O II] $\lambda 7330 \text{ \AA}$	7330 \AA

we perform another fit using a different library, described in the next section.

Finally, we corrected the spectra for internal gas extinction using the Calzetti (2001) attenuation law. The extinction constants $E(B - V) = 0.44 \times 6.60 \log(I_{H\alpha}/I_{H\beta}/2.87)/4.04$ were estimated from the $H\alpha/H\beta$ Balmer decrements following Calzetti (2001), adopting intrinsic ratios $I_{H\alpha}/I_{H\beta} = 2.87$ (Osterbrock 1989).

The factor of 0.44 stems from the correction for differential extinction between the stellar populations and the gas following Calzetti, Kinney & Storchi-Bergmann (1994). As our sample consists of $H II$ galaxies only, which is a subclass of starburst galaxies, using this factor is justified. The emission line strengths were dereddened as well, using the same $E(B - V)$ values but without multiplying them by 0.44.

To make sure our sample contains only $H II$ galaxies, we limited it to galaxies with strong emission lines at a high excitation level, by applying the two criteria:

$$\log\left(\frac{[\text{O III}]\lambda 5007 \text{\AA}}{H\beta}\right) \geq 0.2 \parallel \log\left(\frac{[\text{N II}]}{H\alpha}\right) \leq -0.5, \quad (1)$$

whereas $[\text{N II}] = [\text{N II}] \lambda 6548 \text{ \AA} + [\text{N II}] \lambda 6584 \text{ \AA}$. Of course, these criteria are more restrictive than the traditional definition of $H II$ galaxies as galaxies showing strong Balmer lines. In fact, as our criteria are designed to identify galaxies with gas at a high excitation level, our sample is biased towards extreme case $H II$ galaxies. As a consequence, we expect the galaxies of our sample to show typical properties of $H II$ galaxies, but at a more pronounced level. For example, it consists of $H II$ galaxies with low mass-to-light ratios and thus contains mainly dwarf galaxies. The reason for this approach, apart from emphasizing on typical $H II$ galaxy properties, is to make sure that no ‘contaminating’ objects, such as low-ionization nuclear emission line regions (LINERs), enter the sample. Then, we separated the star-forming galaxies from active galactic nuclei (AGNs), using the Kewley et al. (2001) excitation line criterion,

$$\log\left(\frac{[\text{O III}]\lambda 5007 \text{\AA}}{H\beta}\right) < \frac{0.61}{\log([\text{N II}]/H\alpha) - 0.47} + 1.19, \quad (2)$$

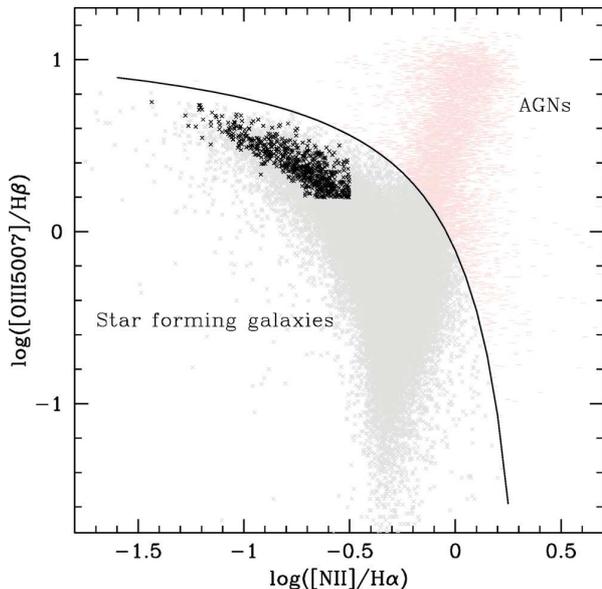


Figure 2. Excitation diagram for all galaxies of our sample (black crosses). The grey crosses represent the test sample, whereas galaxies identified as AGNs are shown as dashes. The solid line shows models from Kewley et al. (2001) used to separate the AGNs from star-forming galaxies.

as shown in Fig. 2.

Our final sample contains 712 H II galaxy spectra. The relatively low number of galaxies compared to the total number of over 900 000 galaxies in the SDSS Data Release 7 is due to our high requirements towards the quality of the spectra, and to the criterion that the spectra must contain all the lines necessary for our study. This way, we guarantee that the sample is not contaminated, and that the obtained results are trustworthy. However, the sample is large enough to be statistically significant.

We also defined a test sample, containing galaxies with some line emission, but not necessarily at a high level, that is H α , H β , [O II] λ 3727 Å, [O III] λ 4959 Å, [O III] λ 5007 Å, [N II] > 0 . The reason for defining a test sample is, on the one hand, to identify properties typical of H II galaxies through comparison with the test sample and, on the other hand, to check our method, by comparing the results for our test sample with the ones obtained in other studies using similar galaxy samples. As we needed to measure the [O II] λ 3727 Å line strength for the test sample as well, it is also limited to galaxies with redshift $\gtrsim 0.033$. To guarantee sufficient quality, we applied the S/N criterion $S/N_g \geq 5$, where the S/N_g values were taken from the SDSS data base. Here too, we removed the AGNs using the Kewley et al. (2001) criterion. This way, our test sample contains 74 989 spectra. Given the criterion that the test sample galaxies show some line emission, it consists of star-forming galaxies, probably mostly spirals.

Fig. 2 shows both our sample of H II galaxies (in black) and the test sample (in grey) in a Baldwin–Phillips–Terlevich (BPT; Baldwin, Phillips & Terlevich 1981) classification diagram.

3 METHOD

3.1 Determination of the gas metallicities

We first measured the gas metallicity, which corresponds to the metallicity of the young stellar populations. In order to derive gas

abundances from its line emission, hydrogen lines are needed, and the lines of at least one ion, generally oxygen, in its two dominant ionization stages. Unfortunately, the electron temperature method using the [O III] λ 4363 Å emission line could not be used as this line and other auroral lines are much too weak to be measured with the necessary precision at the S/N values of the SDSS spectra. In most spectra, we do not even detect the [O III] λ 4363 Å line at all. Hence, we are restricted to strong line methods. The [O II] λ 7330 Å line could be used, but it is too sensitive to the electron temperature and is possibly ‘contaminated’ by recombination contributions.

We compare the gas metallicities derived from 11 different strong line methods.

A frequently used indicator is the [N II] λ 6548 Å or the [N II] λ 6584 Å line (Denicoló, Terlevich & Terlevich 2002; Pettini & Pagel 2004). However, since nitrogen is produced and destroyed during both primary and secondary nucleosynthesis, its abundance correlates in a non-evident way on (oxygen) metallicity. Therefore, it should be used only for rough estimates, to discriminate between the different branches of a multiple-valued method, for instance.

One such multiple-valued indicator is the R_{23} parameter (Pagel et al. 1979), which sums up the fluxes of the [O II] λ 3727 Å and two strong [O III] lines, at 4959 and 5007 Å. As R_{23} is approximately proportional to the oxygen abundance at low metallicities, but decreases at high metallicities due to cooling, this indicator has a low- and a high-metallicity branch. In spite of this double valuedness, the R_{23} parameter is used in many metallicity calibrations (McGaugh 1991; Zaritsky et al. 1994; Pilyugin 2001a,b; Kewley & Dopita 2002; Kobulnicky & Kewley 2004; Pilyugin & Thuan 2005), some of which include a recipe, how to determine, on which branch a given galaxy lies, whereas others leave it up to the user to decide this. In one case, Zaritsky et al. (1994), a calibration is determined only for the high-metallicity branch. Another point that has to be taken into account (although in some methods this is ignored, i.e. Zaritsky et al. 1994) is that R_{23} depends not only on the metallicity of the ionized gas, but also on the hardness of the ionizing radiation, which can be quantified by the ionization parameter q , representing the ionizing photon flux per unit area divided by the number density of hydrogen atoms. Kewley & Dopita (2002) and Kobulnicky & Kewley (2004, ‘new parametrization’ and ‘best estimate’ methods) use the [O III]/[O II] ratio to calculate q in an iterative way from theory, and then determine [O/H] from R_{23} and q , whereas McGaugh (1991) uses the [O III]/[O II] ratio directly in an empirical calibration without determining q . In Pilyugin (2001a,b) and Pilyugin & Thuan (2005), the so-called excitation or ionization or, simply, P parameter is introduced for their empirical calibrations. It is defined as the [O III]/([O II]+[O III]) ratio, so the [O III]/[O II] ratio is related to it, [O III]/[O II] = $P/(1 - P)$. Recapitulating, in all of these methods the gas metallicity is calculated from the R_{23} parameter and the [O III]/[O II] ratio.

Other strong line methods include Kobulnicky & Kewley (2004, [N II] method), combining the [N II] λ 6584 Å line and the [O III]/[O II] ratio, and Pettini & Pagel (2004), employing the ratio between the [O III] λ 5007 Å and [N II] λ 6584 Å lines.

Fig. 3 shows the metallicities of our H II galaxy sample as determined by nine of the methods we studied, plotted against the ones determined by the 10th method, Pilyugin & Thuan (2005), which is the one we ended up adopting. The only calibration not shown here due to space limitations is the Pettini & Pagel (2004) [N II] one, which is a vertically squeezed and shifted downward version of the Denicoló et al. (2002) calibration (first panel).

Most methods yield similar relative metallicities, i.e. the methods agree upon the metallicity differences between galaxies. However,

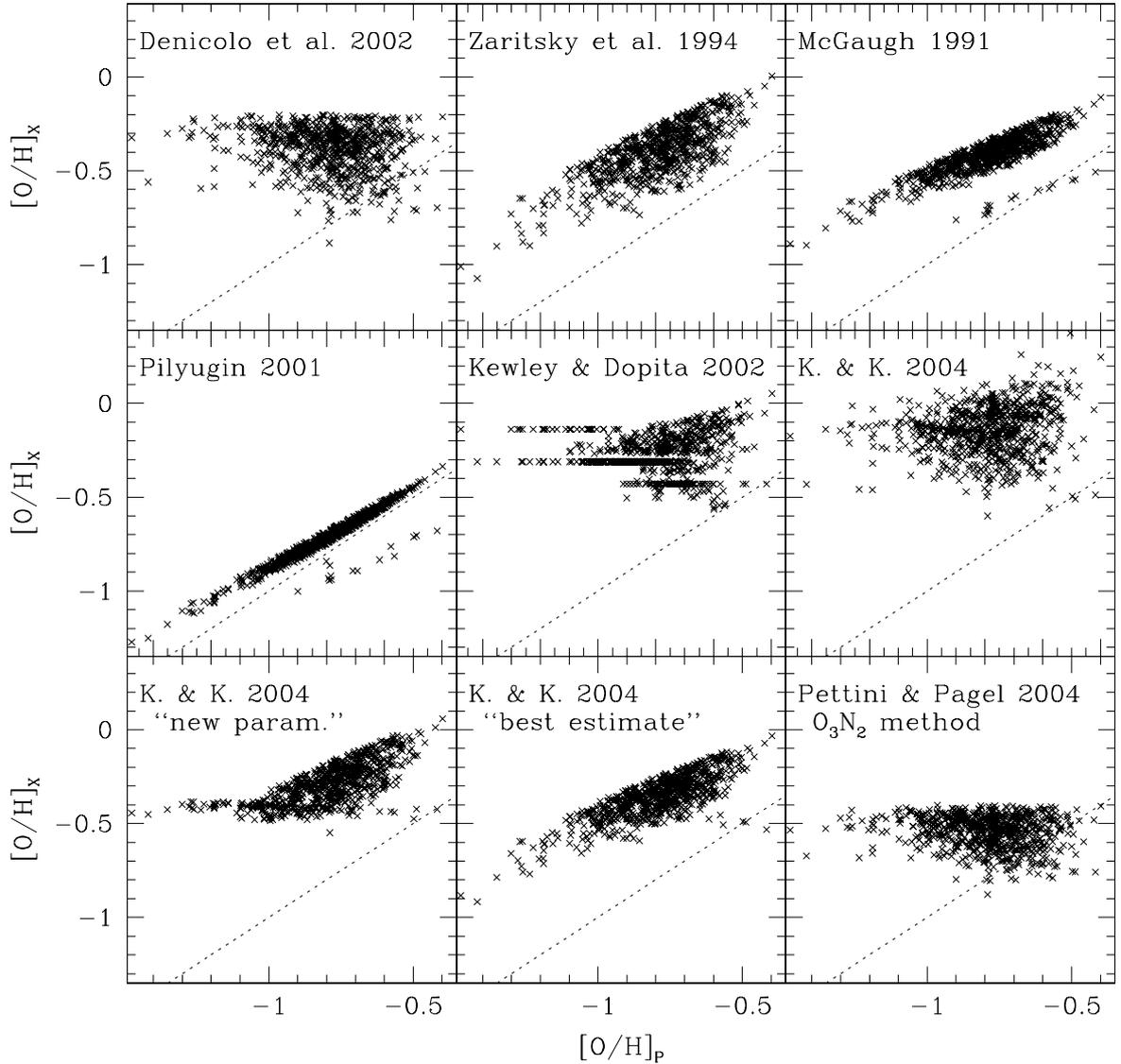


Figure 3. Comparison between the metallicities determined using the Pilyugin & Thuan (2005) method (x -axes) and the ones determined by nine other strong-line methods (y -axes). K. & K. 2004 stands for Kobulnicky & Kewley (2004). The dashed lines show unity, $[O/H]_x = [O/H]_p$. For further details, see text.

when it comes to absolute $[O/H]$ values, systematic differences of up to 0.5 dex between methods can be observed. All methods apart from Pilyugin (2001a,b) and Pilyugin & Thuan (2005, x -axis) result in unrealistically high gas metallicities for $H\text{II}$ galaxies. Especially the Kewley & Dopita (2002) and Kobulnicky & Kewley (2004) ones hardly ever give metallicities below -0.5 . On top of that, the Kewley & Dopita (2002) and Kobulnicky & Kewley (2004, $[N\text{II}]$ and ‘new parametrization’) methods did not always converge. The horizontal lines in the Kewley & Dopita (2002) panel are due to some cases, in which the expression $\log(O/H)+12 = \frac{-k_1 + \sqrt{k_1^2 - 4k_2(k_0 - R_{23})}}{2k_2}$ gave complex values. In these cases, we used only the real part, $-k_1/2k_2$, which results in identical metallicity values for galaxies within the same ionization parameter range. When ignoring these galaxies, we obtain a similar relation as for the Kobulnicky & Kewley (2004) ‘new parametrization of the Kewley & Dopita (2002)’ method.

The method that yields the most realistic metallicities for the galaxies of our sample is the one by Pilyugin & Thuan (2005). It also accounts for the hardness of the ionizing radiation by means of the P parameter. Therefore, we adopted it for our gas metallicity determination. To determine on which branch a given galaxy lies, we used the metallicity estimate based on the $[N\text{II}]\lambda 6548\text{\AA}/H\alpha$ ratio by Denicoló et al. (2002), as suggested by various authors (Pérez-Montero & Díaz 2005; Hoyos et al. 2007, and others). Most galaxies of our sample lie on the high-metallicity branch.

López-Sánchez & Esteban (2010) come to the same conclusion after comparing a series of empirical methods, like the ones presented in this section, with the results of the electron temperature method for a sample of 31 Wolf–Rayet galaxies. They conclude that the nowadays best suitable method for star-forming galaxies, where auroral lines such as $[O\text{III}]\lambda 4363\text{\AA}$ are not observed, is Pilyugin

& Thuan (2005), and that other methods based on photoionization models yield metallicities systematically 0.2–0.3 dex higher and higher dispersion than the Pilyugin & Thuan (2005) calibration.

3.2 Determination of the stellar metallicities

Then, we determined the metallicity of the stellar content, i.e. of the stellar populations making up the galaxies, using a population synthesis method, described in Cuisinier et al. (2006) and Lisker et al. (2006), based on the full spectral fitting in the 3820–8570 Å range, whereas the upper limit can be lower than 8570 Å (in some rare cases as low as 7200 Å), where the deredshifted SDSS spectra do not extend up to this value.

As opposed to the method used for the determination of the line intensities, for this population synthesis we used SSP spectra from the so-called ‘BC99’ SSP library (Westera et al. 2004; Cuisinier et al. 2006; Lisker et al. 2006). It was produced using the GISEL code, implementing the Padova 2000 isochrones (Girardi et al. 2000) combined with the BaSeL 3.1 ‘Padova 2000’ stellar library (Westera 2001; Westera et al. 2002). The BaSeL 3.1 library was calibrated to reproduce the spectral shapes of stars of metallicities [Fe/H] from –2 to 0, so it is the ideal choice for full spectral fitting. By contrast, the ‘BC03’ library described in the previous section would not have been a good choice, for the reasons mentioned there. As the GISEL spectra do not include nebular continuum emission, we added it to the spectra, in the same way as described in Westera et al. (2004), Cuisinier et al. (2006) and Lisker et al. (2006).

In Westera et al. (2004), we found that the stellar content of H II galaxies is made up of at least three populations, a young one (up to 10 Myr old) that is responsible for the ionization of the gas and, therefore, for the nebular emission, an intermediate one (from 20 to 500 Myr), and an old one (at least 1 Gyr old). We showed that these three populations are necessary – and sufficient – to characterize an H II galaxy, which is now the current view (Hoyos et al. 2007). Therefore, in this work we modelled the actual population as being composed of an old, an intermediate and a young stellar population. In a full spectral fit, we determined the masses, ages and metallicities of these partial populations. To obtain meaningful results, we chose to reduce the number of fitting parameters (and thus of degeneracies) to a minimum.

As the young population should be more or less coeval with the gas, whose metallicity is the one attained now by these galaxies, we assumed the young and intermediate populations to have the same chemical composition as the gas. As the gas metallicities we measured are in terms of [O/H], whereas our stellar population library was calibrated in [Fe/H], we still had to translate the oxygen into iron abundances, since the element ratios in H II galaxies are far from solar. Oxygen is released by stellar winds and by SNe I and II, whereas iron and other iron-peak elements are only produced in SNe Ia, which appear only around 1 Gyr after the formation of a stellar population. Therefore, the ratio between the abundances of these two elements in a galaxy depends on its entire star formation history. Adding to this the possibility of ‘selective mass loss’, as is predicted by the models mentioned in the introduction, it would be illusory to expect a unique iron-to-oxygen ratio for the galaxies of our sample. However, as our sample is very homogeneous, we do expect there to exist a mean relation and only a moderate scatter around this relation. Unfortunately, transformations between the two metallicity indicators for H II galaxies are rare in the literature. In earlier abundance studies of this type of galaxies, such as Kniazev et al. (2004) and Izotov et al. (2006), the iron abundance is not determined, as they are also based on emission lines in SDSS spectra.

On the other hand, most theoretical transformations, like the ones determined for the Milky Way bulge and the solar neighbourhood by Matteucci et al. (1999), are not valid for H II galaxies. Finally, we derived the [O/Fe] to [Fe/H] relation from the one for irregular galaxies from Calura et al. (2009), who calculated the element ratios of galaxies of different morphological types using chemical evolution models. According to their fig. 13, the [O/Fe] to [Fe/H] relation for irregular galaxies is around $[O/Fe] = -0.25 \times [Fe/H] - 0.425 \pm 0.15$ independent of redshift, which translates to $[Fe/H] = 1.333 \times [O/H] + 0.5666 \pm 0.2$. Hence, we fixed the metallicities of the young and intermediate populations, $[Fe/H]_y$ and $[Fe/H]_i$, to $[Fe/H]_y = [Fe/H]_i = 1.333 \times [O/H]_p + 0.5666$. (3)

For the galaxies of our sample, most of which have $[O/H]_p$ values in the range from –1 to –0.5, we obtain [O/Fe] ratios between –0.2 and –0.4. As opposed to the metallicity of the young and intermediate populations, the metallicity of the old population, $[Fe/H]_o$, is a free parameter of our fitting procedure, limited to the metallicities of the spectra of the ‘BC99’ SSP library, –2.252, –1.65, –1.25, –0.65, –0.35, 0.027 and 0.225. Contrary to our approach in previous works, we did not force the metallicities of the old populations to be lower than the ones of the younger stars, since one of the aims of this work is to verify if they really are.

As the decomposition of galaxy spectra into SSP spectra is known to present various degeneracies, we verified if our fitting procedure is able to recover the population parameters of synthetic composite spectra made up of three SSPs with noise added. We found that the masses, ages and metallicities of the input populations were properly recovered for the S/N values of both our H II galaxy sample and our test sample (≥ 8 and 5 respectively in the *g* band), and conclude that the number of free parameters of our procedure is adequate and that the results presented in the following can be trusted to be meaningful.

In order to determine the total stellar masses of our galaxies, we had to know their luminosity distances D_L , and their (stellar) mass-to-light ratios. We calculated the former from the redshifts given by the SDSS, using a cosmology of $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. The stellar mass-to-light ratios result from our best fits.

4 RESULTS

The strengths of the emission lines used for this analysis, as well as the internal reddening values $E(B - V)$, the gas metallicities $[O/H]_p$ and the total stellar masses, M_{tot} , can be found in Table A1 with the electronic version of this paper (see Supporting Information).

Fig. 4 shows the stellar mass and gas metallicity distributions of our sample (in black) and the test sample (in grey). The average of the logarithms of the stellar masses (in M_\odot) amounts to 9.4 and the average gas metallicity to –0.79, whereas those values are 9.8 and –0.78 respectively for the test sample.

As expected, the galaxies of the H II galaxy sample have, on average, lower masses than the ones of the test sample. Since our sample contains very few galaxies with stellar masses below $2 \times 10^8 M_\odot$ or above $2 \times 10^{10} M_\odot$ (the mass bins at 5×10^6 , 1×10^7 , 2×10^7 , 5×10^7 , 1×10^8 ; 5×10^{10} , 1×10^{11} , 2×10^{11} , 5×10^{11} and $1 \times 10^{12} M_\odot$ contain 1, 0, 4, 2, 3; 1, 2, 0, 0 and 0 galaxies, respectively), statistical studies of galaxies in these mass ranges suffer from low number statistics. In the following we will only consider the range from 2×10^8 to $2 \times 10^{10} M_\odot$ when studying galaxy properties as a function of mass. In Figs 7 and 8, this mass range is delimited by dotted lines. For the test sample, the mass

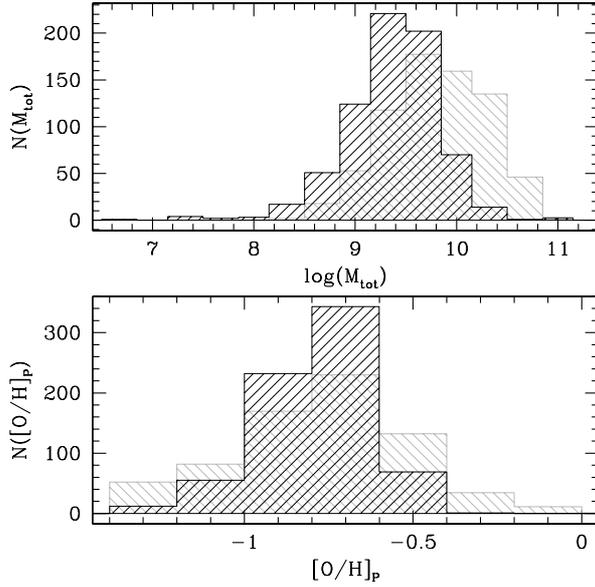


Figure 4. Stellar mass (top panel) and gas metallicity (bottom panel) distributions of the galaxies of our sample (in black). The mass and gas metallicity distributions of the test sample are shown in grey (scaled to the same total number of galaxies).

range from 2×10^8 to $1 \times 10^{11} M_{\odot}$ could be used for such studies, as each mass bin contains over 150 galaxies.

The average gas metallicities are nearly the same for both samples, but the distribution is narrower for the H II galaxies, indicating that this is a more homogeneous sample.

Fig. 5 shows the metallicity distributions of the partial populations. As we fixed $[\text{Fe}/\text{H}]_y = [\text{Fe}/\text{H}]_i = 1.333 \times [\text{O}/\text{H}]_p + 0.5666$

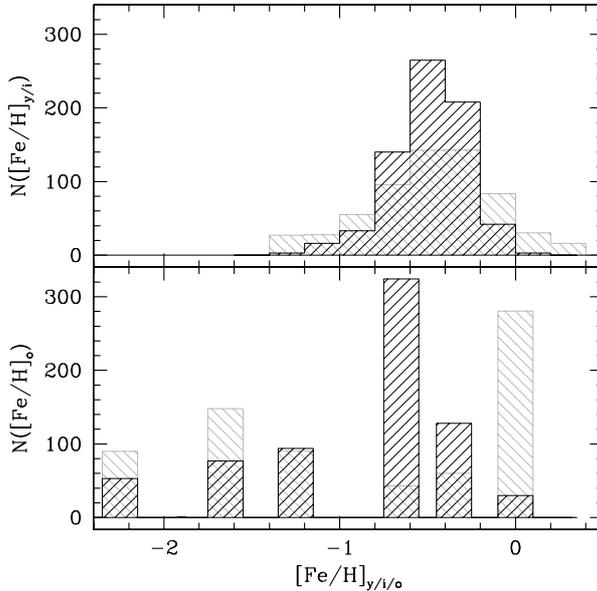


Figure 5. Metallicity distributions of the partial populations in our galaxy sample. Top panel: the joint young- and intermediate-age populations. Bottom panel: old populations. The metallicity distributions of the partial populations in the test sample are shown in grey (scaled to the same total number of galaxies).

in the fitting procedure, we show in the upper panel the metallicity of the joint young- and intermediate-age population. The average metallicity of the joint young and intermediate populations is -0.49 , and of the old one -0.87 (test sample: -0.49 and -0.84). Here too, the distributions are narrower for the more homogeneous H II galaxies sample. Nevertheless, the spread in $[\text{Fe}/\text{H}]_o$ is wide. As $[\text{Fe}/\text{H}]_o$ is a measure of the average metallicity of all populations older than ~ 1 Gyr, this wide spread reflects the various stages of (metallicity) evolution of the galaxies of our sample around 1 Gyr ago.

5 DISCUSSION

In Fig. 6, we show the gas metallicity in function of the metallicity of the old population, both for the H II galaxies and for the test sample. In either sample, not much of a relation can be seen, any possible trend being significantly smaller than the scatter. Apparently, the stage of a galaxy's evolution around 1 Gyr ago is not necessarily reflected in the present-day one. The chemical evolution and the present-day (gas) metallicities of galaxies suffer stronger influences from other factors, such as the galaxy mass.

Fig. 7 shows how the present gas metallicity $[\text{O}/\text{H}]_p$ depends on the total stellar galaxy mass. In the test sample, $[\text{O}/\text{H}]_p$ clearly increases with mass in the range from 2×10^8 to $1 \times 10^{11} M_{\odot}$, as has been found in previous studies (Tremonti et al. 2004; Thomas et al. 2005; Asari et al. 2007; Panter et al. 2008, and others). However, our mass-metallicity relation lies lower than the ones found in these works, as recent studies (Ellison et al. 2008; Mannucci et al. 2010) suggest to be the case for samples of star-forming galaxies, such as our test sample.

The H II galaxy sample represents galaxies of a more homogeneous class than the test sample. Therefore, in this sample, the mass-metallicity relation is much less pronounced, if existing at all. In fact, the average gas metallicity stays constant between -0.7 and -0.8 in the full galaxy mass range from 2×10^8 to $2 \times 10^{10} M_{\odot}$. Panter et al. (2008) find almost the same relation for their subsample

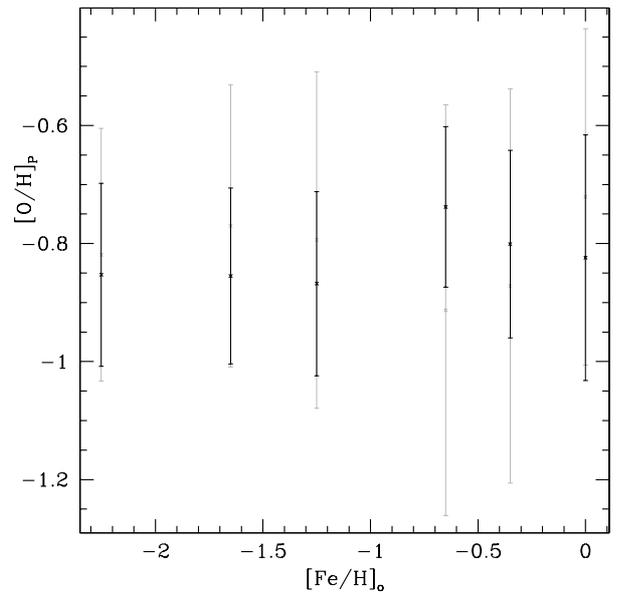


Figure 6. Gas metallicities as a function of the metallicities of the old populations. The crosses represent the average of the gas metallicities for each old population metallicity value, and the error bars show the standard deviations.

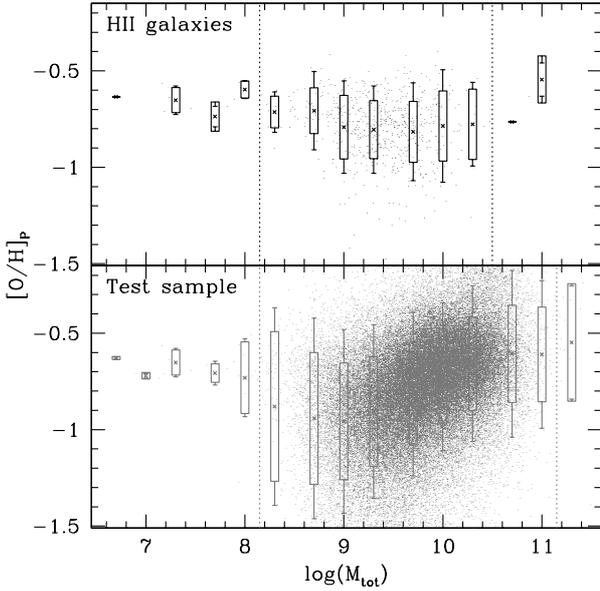


Figure 7. Top panel: the gas metallicity $[O/H]_P$ as a function of the total stellar mass of the galaxies of our sample. The dots show the individual galaxies, whereas the crosses, boxes and bars show the statistical properties after subdividing the sample into mass bins. For each value of M_{tot} , the crosses represent the average, the boxes represent the standard deviation and the vertical bars represent the 90th percentile range. The dotted vertical lines detach the ‘trustworthy’ range from the nearly empty mass bins. Bottom panel: the individual galaxies, average, standard deviation and 90th percentile range of $[O/H]_P$ for the test sample.

of galaxies that are (flux-)dominated by young populations (≤ 0.5 Gyr), as can be seen in their fig. 10.

In a ‘closed box’ context, this would mean that the gas fraction that has been transformed into stars so far is in the same range for all galaxies of the H II galaxy sample, independent of their masses. However, H II galaxies are not ‘closed boxes’, and our results should be interpreted in the light of models including interaction processes with the environment, like the ones presented in Section 1.

Most of these models predict that H II galaxies lose a large fraction of their heavy elements through galactic winds, whose escape efficiencies depend upon the galaxy mass. These models are at odds with our result for two reasons: they result in a dependence of the metallicity on the galaxy mass and their predicted present-day metallicities are lower than the ones we measure.

The Recchi et al. (2006) and Recchi & Hensler (2007) models with constantly infalling clouds appear more promising, especially the models NGC-4+BC of Recchi et al. (2006) and NCSM of Recchi & Hensler (2007). The clouds partially retain the metal-rich gas, resulting in higher present-day metallicities. Depending on the star formation history and on the mass of the infalling clouds, galaxies like the ones described in these models can reach the metallicities we measure. On top of that, this scenario provides a possible explanation, why we find metallicities in the same ranges for different stellar masses. Not only the SFR and, therefore, the amount of produced metals and galactic wind increase with increasing galaxy mass, but also the amount of infalling low-metallicity clouds and of retained outflowing high-metallicity gas. Of course, only detailed model calculations can tell, if the mass dependencies of the different metallicity-increasing and -decreasing processes really add up

in such a way, that the final metallicity becomes mass-independent. The fact that a constant mass–metallicity relation is observed only in our H II galaxy sample and not, for example, in our test sample, or in any of the studies mentioned at the beginning of this section, shows that only in galaxies with extreme star formation, and thus extreme stellar winds, the metal outflow manages to balance the re-entention in this way. However, the models including infalling clouds seem to be on the right track.

Another noteworthy point of Fig. 7 is that though the average gas metallicity $[O/H]_P$ does not vary with the galaxy mass, its dispersion does increase with galactic mass. This indicates that high-mass galaxies have a more complex chemical, and certainly assembly, history.

Apart from the present-day metallicity, our analysis also allows us to quantify the chemical evolution of our sample galaxies since the formation of the old population. As the metallicity of the young (and intermediate) population represents the galaxy’s present-day metallicity, and the one of the old population is a measure of the average metallicity of all stars older than ~ 1 Gyr, the quantity $[Fe/H]_{y/i} - [Fe/H]_o$, which we shall call the metal enrichment, is a good measure of the change in metallicity during the last few Gyr. Fig. 8 shows the metal enrichment in function of the stellar mass for the galaxies in both our H II galaxies sample and the test sample. For our H II galaxies, the enrichment is fairly constant of the order of 0.3 up to $5 \times 10^9 M_{\odot}$, and almost 0.5 dex higher for the last two mass bins within the mass range, where reliable statistics are possible, 1×10^{10} and $2 \times 10^{10} M_{\odot}$. This suggests that the enrichment mechanism is independent of mass up to $5 \times 10^9 M_{\odot}$, which is compatible with the formation scenario favoured in the previous paragraph. For galaxy masses above $5 \times 10^9 M_{\odot}$, the enrichment is significantly higher, which means that at high

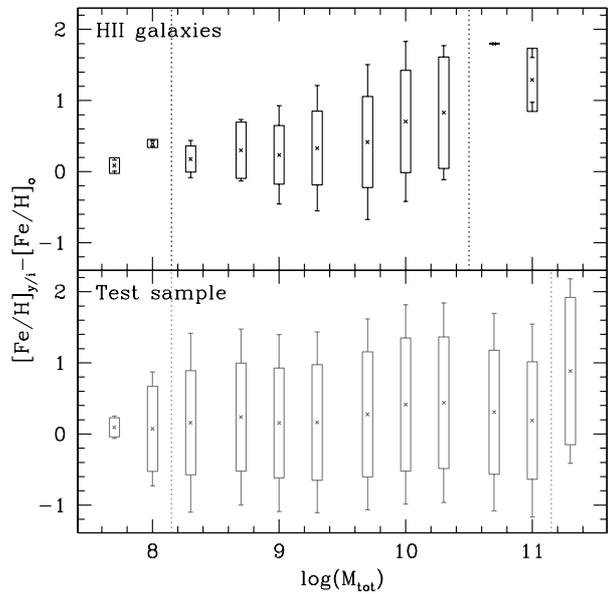


Figure 8. Top panel: the metal enrichment, $[Fe/H]_{y/i} - [Fe/H]_o$, as a function of the total mass of the galaxies of our sample. For each value of M_{tot} , the crosses represent the average, the boxes represent the standard deviation and the vertical bars represent the 90th percentile range. The dotted vertical lines detach the ‘trustworthy’ range from the nearly empty mass bins. Bottom panel: the average, standard deviation and 90th percentile range of $[Fe/H]_{y/i} - [Fe/H]_o$ for the test sample.

masses, the metal-rich gas could not escape the galaxy's potential well, as predicted in the MacLow & Ferrara (1999) models. Here too, the dispersion increases with galaxy mass, confirming more complex chemical and assembly histories for high-mass galaxies.

An interesting point is that we find a negative metal enrichment, thus a metal reduction, for 164 galaxies, i.e. 23 per cent of our sample. Two mechanisms are known to reduce the (gas) metallicity of a galaxy: selective outflow of metal-enriched gas and infall of low-metallicity gas clouds, maybe even of primordial composition. Both processes might have contributed to the metal reduction of our galaxies, but in the light of our previous results, we expect the second one to be more important on long time-scales.

It is intriguing that in the test sample, too, we find the metal enrichment to be mass-independent, even though the final metallicity increases with galaxy mass. $[\text{Fe}/\text{H}]_{y/i} - [\text{Fe}/\text{H}]_o$ amounts to around 0.36 in the full range from 2×10^8 to $1 \times 10^{11} M_{\odot}$. Even the dispersion is fairly constant over this mass range. Apparently, enrichment mechanisms are similar for galaxies of different masses even in a heterogeneous sample like our test sample. Around 44 per cent of the test sample galaxies, 33 178 in total, show negative metal enrichment, which proves that processes like selective outflow or low-metallicity gas infall are also at play in these galaxies.

6 CONCLUSIONS

We performed a gas and stellar population analysis of H II galaxies from the SDSS Data Release 7, using their full spectra. We selected a sample which contains 712 H II galaxy spectra, a relatively low number due to our high requirements towards the quality of the spectra and to the criterion that the spectra contain all the spectral lines necessary for our study. We also selected a test sample of 74 989 galaxies, using less restrictive requirements, the only criterion being that they show some line emission. Therewith, it is a sample of star-forming galaxies.

We derived independent metallicities for the young populations present in our sample galaxies from the gas emission lines, and for the old populations from full spectra fitting, i.e. from the continuum and absorption features from stars.

Our test sample follows the well-known mass–metallicity relation, e.g. the fact that the (gas) metallicity increases with (stellar) galaxy mass, indicating that low-mass galaxies are chemically less evolved than high-mass galaxies. We do not find any systematical tendency of the difference between the metallicities of the young and the old stellar populations, a quantity we call metallicity enrichment, with galactic mass, indicating that, on average, recycling mechanisms of the interstellar medium should be the same for low- and high-mass galaxies.

In our H II galaxy sample, on the other hand, we find no mass dependence of the present-day metallicity. For the metallicity enrichment, we find no mass dependence for masses up to $5 \times 10^9 M_{\odot}$, whereas above this mass, the enrichment is significantly higher.

We interpret these findings in the light of recent hydrodynamical evolutionary models of dwarf irregular galaxies, of which H II galaxies are a subclass. We favour models featuring constantly infalling low-metallicity clouds able to retain part of the high-metallicity galactic winds, such as the models NGC-4+BC from Recchi et al. (2006) and NCSM from Recchi & Hensler (2007), since models without infalling clouds fail to predict the metallicities and the non-dependence on galaxy mass of our sample galaxies. For galaxies with masses above $5 \times 10^9 M_{\odot}$, most, if not all, of the metal-rich gas is retained by the galaxy's gravitational potential, as predicted

by MacLow & Ferrara (1999), which leads to the observed breakdown of the constant mass–enrichment relation.

However, the dispersion of both the present-day metallicity and the metallicity enrichment does increase with galactic mass, indicating more complex chemical and assembly histories for high-mass H II galaxies.

In 23 per cent of our H II galaxies, the metallicity has decreased in the last few Gyr. We attribute the metallicity decrease mainly to the dilution of the galactic gas by the infalling clouds, which is compatible with the models we favour.

ACKNOWLEDGMENTS

We would like to thank the Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and the PCI programme of ON/MCT (DTI/CNPq) for financial support. We would like to thank as well Eduardo Telles and Helio Rocha-Pinto for fruitful discussions. Finally, the authors acknowledge support by the Swiss National Science Foundation.

REFERENCES

- Abazajian K. N. et al., 2009, *ApJS*, 182, 543
 Asari N. V., Cid Fernandes R., Stasińska G., Torres-Papaqui J. P., Mateus A., Sodré L., Schoenell W., Gomes J. M., 2007, *MNRAS*, 381, 263
 Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, 93, 5
 Bruzual A. G., Charlot S., 1993, *ApJ*, 405, 538
 Bruzual A. G., Charlot S., 2003, *MNRAS*, 344, 1000
 Calura F., Pipino A., Chiappini C., Matteucci F., Maiolino R., 2009, *A&A*, 504, 373
 Calzetti D., 2001, *PASP*, 113, 1449
 Calzetti D., Kinney A. L., Storchi-Bergmann T., 1994, *ApJ*, 429, 582
 Charlot S., Bruzual A. G., 1991, *ApJ*, 367, 126
 Cid Fernandes R., Leão J. R. S., Lacerda R. R., 2003, *MNRAS*, 340, 29
 Cid Fernandes R., Asari N. V., Sodré L., Jr, Stasińska G., Mateus A., Torres-Papaqui J. P., Schoenell W., 2007, *MNRAS*, 375, L16
 Cuisinier F., Westera P., Telles E., Buser R., 2006, *A&A*, 423, 133
 Denicoló G., Terlevich R., Terlevich E., 2002, *MNRAS*, 329, 315
 Ellism S. L., Patton D. R., Simard L., McConnachie A. W., 2008, *ApJ*, 672, L107
 Fagotto F., Bressan A., Bertelli G., Chiosi C., 1994, *A&AS*, 105, 39
 Fitzpatrick E. L., 1999, *PASP*, 111, 63
 French H. B., 1980, *ApJ*, 240, 41
 Girardi L., Bressan A., Chiosi C., Bertelli G., Nasi E., 1996, *A&AS*, 117, 113
 Girardi L., Bressan A., Bertelli G., Chiosi C., 2000, *A&AS*, 141, 371
 Hägele G. F., Díaz A. I., Terlevich E., Terlevich R., Pérez-Montero E., Cardaci M. V., 2008, *MNRAS*, 383, 209
 Hoyos C., Guzmán R., Díaz A. I., Koo D. C., Bershady M. A., 2007, *AJ*, 134, 2455
 Izotov Y. I., Stasińska G., Meynet G., Guseva N. G., Thuan T. X., 2006, *A&A*, 448, 955
 Kehrig C., Telles E., Cuisinier F., 2004, *AJ*, 128, 1141
 Kehrig C., Vilchez J. M., Telles E., Cuisinier F., Pérez-Montero E., 2006, *A&A*, 457, 477
 Kewley L. J., Dopita M. A., 2002, *ApJS*, 142, 35
 Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, *ApJ*, 556, 121
 Kniazev A. Y., Pustilnik S. A., Grebel E. K., Lee H., Pramskij A. G., 2004, *ApJS*, 153, 429
 Kobulnicky H. A., Kewley L. J., 2004, *ApJ*, 617, 240
 Kong X., Charlot S., Weiss A., Cheng F., 2003, *A&A*, 403, 877
 Le Borgne J.-F. et al., 2003, *A&A*, 402, 433
 Lequeux J., Peimbert M., Rayo J. F., Serrano A., Torres-Peimbert S., 1979, *A&A*, 80, 155
 Lisker T., Glatt K., Westera P., Grebel E. K., 2006, *AJ*, 132, 2432

López-Sánchez Á. R., Esteban C., 2010, *A&A*, 517, A85
 McLaugh S. S., 1991, *ApJ*, 380, 140
 MacLow M.-M., Ferrara A., 1999, *ApJ*, 513, 142
 Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, *MNRAS*, 408, 2115
 Matteucci F., Romano D., Molaro P., 1999, *A&A*, 341, 458
 Osterbrock D. E., 1989, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei*. University Science Books, Mill Valley, CA
 Pagel B. E. J., Edmunds M. G., Blackwell D. E., Chun M. S., Smith G., 1979, *MNRAS*, 189, 95
 Panter B., Jimenez R., Heavens A. F., Charlo S., 2008, *MNRAS*, 391, 1117
 Pérez-Montero E., Díaz A. I., 2003, *MNRAS*, 346, 105
 Pérez-Montero E., Díaz A. I., 2005, *MNRAS*, 361, 1063
 Pérez-Montero E., García-Benito R., Hägele G. F., Díaz A. I., 2010, *MNRAS*, 404, 2037
 Pettini M., Pagel B. E. J., 2004, *MNRAS*, 348, 59
 Pilyugin L. S., 2001a, *A&A*, 369, 594
 Pilyugin L. S., 2001b, *A&A*, 374, 412
 Pilyugin L. S., Thuan T. X., 2005, *ApJ*, 631, 231
 Raimann D., Bica E., Storchi-Bergmann T., Melnick J., Schmitt H., 2000, *MNRAS*, 314, 295
 Recchi S., Hensler G., 2007, *A&A*, 476, 841
 Recchi S., Matteucci F., D’Ercole A., 2001, *MNRAS*, 322, 800
 Recchi S., Matteucci F., D’Ercole A., Tosi M., 2002, *A&A*, 384, 799
 Recchi S., Matteucci F., D’Ercole A., Tosi M., 2004, *A&A*, 426, 37
 Recchi S., Hensler G., Angeretti L., Matteucci F., 2006, *A&A*, 445, 875
 Sargent W. L. W., Searle L., 1970, *ApJ*, 162, L155
 Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525
 Searle L., Sargent W. L. W., Bagnuolo W. G., 1973, *AJ*, 179, 427
 Tenorio-Tagle G., Muñoz-Tuñón C., Pérez E., Silich S., Telles E., 2006, *ApJ*, 643, 186

Terlevich R., Melnick J., Masegosa J., Moles M., Copetti M. V. F., 1991, *A&AS*, 91, 285
 Thomas D., Maraston C., Bender R., Mendes de Oliveira C., 2005, *ApJ*, 621, 673
 Tremonti C. A. et al., 2004, *ApJ*, 613, 898
 Vílchez J. M., Iglesias-Páramo J., 2003, *ApJS*, 145, 225
 Westera P., 2001, PhD thesis, Univ. Basel
 Westera P., Lejeune T., Buser R., Cuisinier F., Bruzual A. G., 2002, *A&A*, 381, 524
 Westera P., Cuisinier F., Telles E., Kehrig C., 2004, *A&A*, 423, 133
 York D. G. et al., 2000, *AJ*, 120, 1579
 Zaritsky D., Kennicutt R. C., Jr, Huchra J. P., 1994, *ApJ*, 420, 87

APPENDIX A: POPULATION PARAMETERS OF INDIVIDUAL SPECTRA

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table A1. Line strengths, internal reddening, gas metallicities and total stellar masses of individual spectra.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Table A1. Line strengths, internal reddening, gas metallicities and total stellar masses of individual spectra (first 10 lines only; for the complete sample, see the Supporting Information with the electronic version of the paper).

SDSS spectroscopic ID MJD-plate-fibre	$I(H\beta)^a$ 4861 Å	$H\alpha^b$ 6563 Å	$[O II]^b$ 3727 Å	$S/N_{[O II]}$ 3727 Å	$[O III]^b$ 4959 Å	$[O III]^b$ 5007 Å	$[N II]^b$ 6548 Å	$[N II]^b$ 6584 Å	$E(B - V)$	$[O/H]_P$	M_{tot} [$10^9 M_{\odot}$]
51637-0306-583	5.102	3.010	5.848	2.948	0.728	2.088	0.148	0.472	0.088	-1.022	0.63
51658-0282-047	10.075	3.103	5.644	2.806	0.694	2.049	0.143	0.457	0.145	-0.991	1.93
51662-0308-628	10.264	3.017	4.800	4.074	0.669	2.049	0.158	0.571	0.093	-0.860	2.11
51663-0307-268	14.076	3.052	4.270	3.005	1.009	3.078	0.176	0.574	0.114	-0.794	3.27
51671-0299-571	8.794	3.179	6.656	2.846	0.615	2.199	0.203	0.626	0.189	-1.142	6.78
51691-0350-439	24.324	3.105	3.533	3.102	0.550	1.696	0.212	0.644	0.146	-0.654	1.98
51692-0339-437	11.115	2.921	2.707	4.272	1.238	3.741	0.074	0.223	0.033	-0.599	0.34
51783-0395-570	9.257	2.924	3.362	4.199	0.639	1.975	0.118	0.374	0.034	-0.624	0.98
51812-0404-507	9.680	3.042	4.195	3.339	0.641	1.964	0.144	0.490	0.108	-0.763	4.10
51818-0383-266	8.743	2.944	3.631	3.398	0.899	2.789	0.097	0.303	0.047	-0.687	0.87

^a $I(H\beta)$ given in units of $10^{15} \text{ erg s}^{-1} \text{ cm}^{-2}$.

^bFlux ratios given in $I(\lambda)/I(H\beta)$.

This paper has been typeset from a \LaTeX file prepared by the author.

Capítulo 7

Conclusão

Acreditamos ter demonstrado que é possível realizar a fotometria superficial de uma grande quantidade de galáxias a partir da manipulação e do tratamento automatizado dos dados. Desenvolvemos os nossos próprios processos e rotinas de redução de dados fotométricos e de computação de índices morfológicos. Como resultado, apresentamos um total de cerca 20 000 linhas de programação, todas escritas em IDL. Uma grande parte da tese é consagrada à descrição técnica dos métodos e algoritmos usados.

Estudamos a morfologia de galáxias, as diferentes maneiras de quantificá-la, a capacidade de implementar classificações automáticas e a possibilidade de se detectar uma assinatura morfológica de interação. Em galáxias HII, mais especificamente, exploramos a interação de galáxias, a formação estelar induzida e a distribuição das regiões de formação estelar. Com dados espectroscópicos, efetuamos uma síntese de população e determinamos a metalicidade, o enriquecimento químico e as massas das populações parciais. Todas as nossas pesquisas foram desenvolvidas em dados extraídos do SDSS.

Podemos destacar os seguintes resultados ou realizações notáveis:

- Desenvolvemos um procedimento próprio, automatizado, de redução de dados fotométricos, especialmente dedicado a galáxias, que consiste em: (1) extração de uma galáxia da imagem de um campo qualquer, (2) detecção dos objetos, (3) remoção dos objetos estranhos à galáxia e (4) subtração do céu. O processo atua sobre as cinco bandas simultaneamente.

7. CONCLUSÃO

- Revisamos as definições de aberturas que podem ser usadas para a fotometria de galáxias. Mostramos que uma abertura baseada em um *threshold* não é conveniente para medidas fotométricas finas, porque depende fortemente da razão sinal/ruído. Aberturas baseadas em uma dimensão específica, tal como os raios de Kron e de Petrosian, devem ser preferencialmente utilizadas. Mostramos também como os segundos momentos da imagem podem ser empregados para melhorar a computação da fotometria de abertura, com a utilização de aberturas elípticas.
- Definimos a “isofota de Petrosian”, uma abertura reunindo as vantagens da isofota e de um raio característico. Em todo o trabalho, a isofota de Petrosian é usada como abertura padrão para todas as medidas fotométricas e morfológicas.
- Para a nossa análise da morfologia quantitativa, calculamos primeiramente os índices habituais: o parâmetro de Sérsic n , o índice de concentração C , o índice de assimetria A , o índice de *clumpiness* S , o coeficiente de Gini G e o índice M_{20} . Tentamos, tanto quanto possível, melhorar e adaptar as definições e os processos apresentados na literatura.
- Definimos o índice de anelipticidade α como uma avaliação da deformação da isofota externa da galáxia. Mais precisamente, α mede se a isofota de Petrosian tem uma forma próxima de uma elipse ou não.
- Definimos o índice de granulação γ como a inclinação do espectro formado pelos momentos radiais de segunda ordem dos coeficientes de *wavelet*. Essa definição, bastante técnica, avalia, de fato, a dispersão radial das estruturas de pequenas escalas. γ pode ser interpretada como um parâmetro que caracteriza se uma galáxia contém regiões de formação estelar fora do seu centro.
- A presença nos dados de muito *outliers* requereram métodos de análise estatística mais robustos. Para isso, adaptamos o coeficiente de correlação por pseudo-covariâncias calculadas com medianas em vez de médias. As correlações entre os índices assim calculados deixam aparecer duas famílias de índices: n , C e G como uma indicação da concentração de luz e A , S e γ como uma indicação da textura da galáxia. M_{20} é uma medida intermediária.

-
- Mostramos que a capacidade de um índice para separar as galáxias *early-type* das *late-type* é melhor para os índices de tipo concentração; mas que a correlação com o tipo morfológico é melhor para os índices de tipo textura, porque ocorre ao longo da totalidade da sequência de Hubble. Isso implica na necessidade de se utilizar os diferentes índices de forma complementar em uma classificação automática.
 - O estudo da robustez estatística dos índices estabeleceu que é a nitidez da imagem, sobretudo, que deprecia a qualidade e o nível de confiança das medidas de fotometria superficial. Definimos, então, um critério de qualidade fotométrica como: diâmetro angular da galáxia ≥ 8 *seeing*.
 - Realizamos quatro classificações morfológicas automáticas para prever o tipo de Hubble. Usando quatro índices morfológicos, G , γ , $\log(A)$ e M_{20} , apresentamos os seguintes algoritmos de classificação: a regressão linear, o modelo de mistura de gaussianas, o modelo de mistura de distribuições empíricas e o método k-médias.
 - Ilustramos o fato de que a escolha de um dado algoritmo não afeta fortemente o sucesso da classificação automática. Usando uma amostra apurada, todos os métodos conseguiram prever o tipo de Hubble com um erro médio de 0,6-0,8, sensivelmente igual ao desvio entre as diferentes classificações visuais. O método ligeiramente melhor é aquele que modela a distribuição de cada tipo com uma gaussiana.
 - Como α mede a deformação da isofota externa, que é consequência de uma perturbação gravitacional, ele pode ser usado para determinar um “indicador de chance de interação”. Associamos α com A , também usado na literatura para identificar interação, para definir: $I.C.I. = A \cdot \alpha$.
 - Aplicamos esse indicador em galáxias HII. As morfologias mais distorcidas são observadas quando as regiões de formação estelar estão distribuídas por toda a área da galáxia (grande γ) e, em média, para galáxias como um grande diâmetro e uma alta taxa de formação estelar específica (sSFR).

7. CONCLUSÃO

- a partir desses resultados, deduzimos que a formação estelar desencadeada pelas perturbações gravitacionais não é máxima quando a galáxia está mais deformada. Se torna máxima em uma fase posterior, quando as instabilidades dinâmicas causam um fluxo de gás significativo para a região central. Durante esta fase final, a formação estelar se localiza essencialmente na região central, mas a morfologia já se normalizou, se tornando menos irregular. Tudo acontece como se a formação estelar respondesse à morfologia perturbada com um atraso.
- Em um estudo espectroscópico de galáxias *late-type*, comparamos onze métodos diferentes de determinação da metalicidade do gás por linhas de emissão. A metalicidade do gás é fixada para a população estelar jovem. Pelo método de síntese de população, determinamos as massas, idades e metalicidades das populações parciais.
- Observamos que uma primeira amostra, formada de galáxias com alta taxa de formação estelar, segue a relação massa-metalicidade padrão, ou seja, a metalicidade do gás aumenta com a massa da galáxia. Isso indica que as galáxias de baixa massa são quimicamente menos evoluídas do que as galáxias de alta massa. Ao contrário, o enriquecimento da metalicidade não mostra dependência sistemática com a massa galáctica. O que traduz o fato de que, em média, o mecanismo de reciclagem do meio interestelar é o mesmo nas galáxias de alta e baixa massas.
- Em uma segunda amostra, formada de galáxias HII selecionadas por critérios muito mais estritos, não constatamos dependência entre a metalicidade do gás e a massa. Com relação ao enriquecimento da metalicidade, reparamos um leve aumento com a massa, mas apenas para massas maiores do que $5 \cdot 10^9 M_{\odot}$. Estas observações favorecem modelos de evolução galáctica com constante queda de gás intra-aglomerado, pobre em metal (*gas infall*). De fato, o gás pobre ao cair impede a expulsão do gás rico, formado na última geração estelar (*gas outflow*). Nas galáxias de alta massa, o potencial gravitacional é suficiente para reter o gás rico em metal e, conseqüentemente, a metalicidade aumenta mais rapidamente.

Apêndice A

Representações gráficas e pontos divergentes

Enquanto intuitivamente parece não haver dúvida de que a representação gráfica permite que uma informação seja mais rapidamente compreendida do que uma tabela de números, a escolha do estilo do gráfico e das grandezas plotadas dependem da situação e podem dar ao leitor uma visão diferente dos resultados. Nesta seção, tentamos explorar várias maneiras de representar uma grande quantidade de dados. Como exemplo, pegamos as 2253 galáxias do catálogo morfológico de Fukugita (Fukugita et al. 2007) e passamos a representar duas grandezas: uma fotométrica, a cor $u-i$, e uma morfológica, o índice de granulação γ .

A informação pertinente não é sempre aquela que chama atenção à primeira vista. O olho, como alguns indicadores estatísticos, atribui uma grande importância aos pontos divergentes (o termo inglês *outliers* é muitas vezes usado). Assim, concentrando-nos na descrição, deveremos aprofundar considerações estatísticas.

Temos, então, 2253 pontos para representar através de três variáveis: $u-i$, γ e o tipo morfológico T . A maneira mais básica de visualizar os dados e de enfatizar uma possível relação entre as variáveis é plotar uma variável contra a outra, para cada ponto. Isto gera os três gráficos da Figura A.1.

O tipo morfológico é uma propriedade intrínseca da galáxia que, possivelmente, define a sua natureza. Queremos saber qual é a relação, se houver uma, entre as medidas físicas $u-i$ e γ para cada tipo morfológico. Vamos, para isso, desenhar a nuvem de

A. REPRESENTAÇÕES GRÁFICAS E PONTOS DIVERGENTES

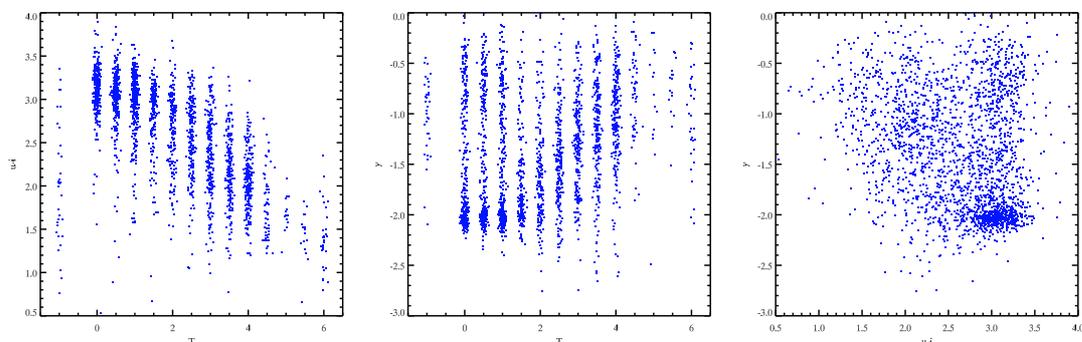


Figura A.1: A três variáveis consideradas ($u-i$, γ e o tipo morfológico T), uma em função da outra. A variável T é artificialmente deslocada para mais clareza.

pontos no plano formado por $u-i$ e γ , usando um código de cores para distinguir entre as diferentes morfologias (Figura A.2).

Em muitos casos, o objeto do estudo não é, propriamente falando, todas as galáxias, mas, sim, as galáxias de um tipo dado. As galáxias elípticas ($T=0$) são tratadas como

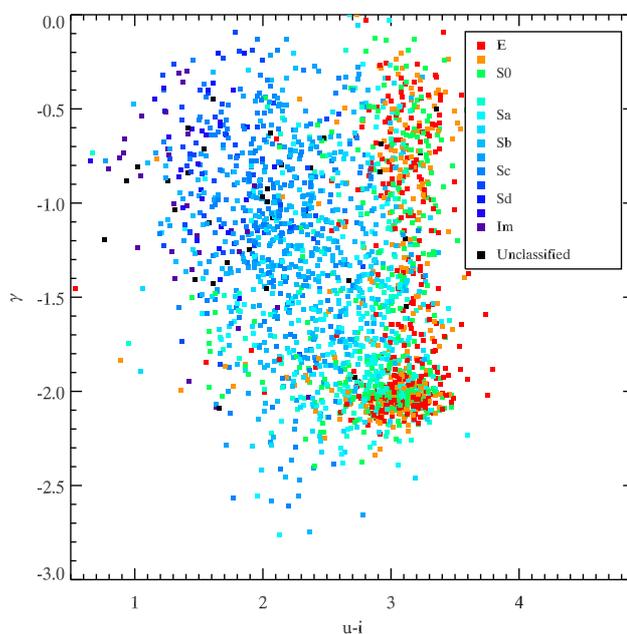


Figura A.2: Nuvem de pontos no plano cor-granulação. A cor dos pontos indica o tipo morfológico.

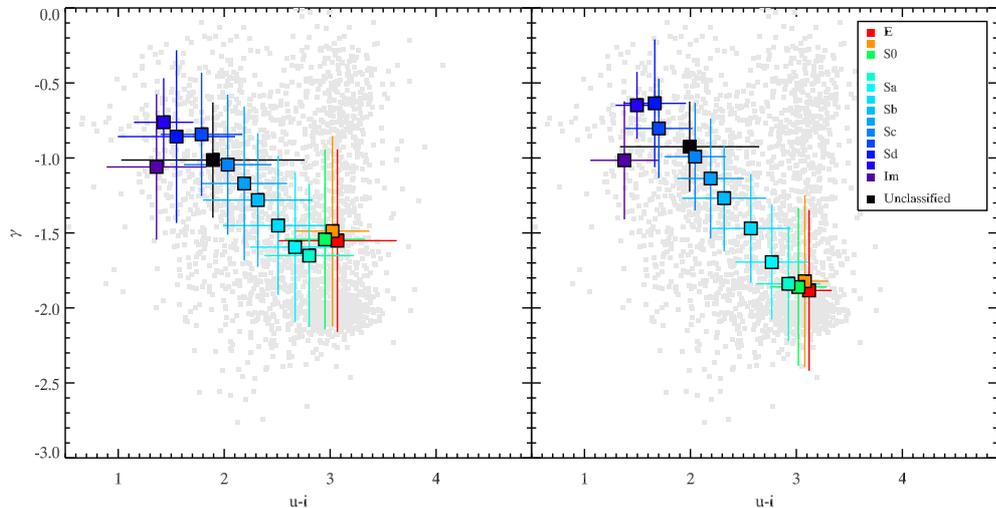


Figura A.3: Plano cor-granulação para os tipos morfológicos. À esquerda: média com barra de 1 desvio padrão; à direita, mediana com barra de 1 desvio absoluto médio.

uma categoria homogênea de galáxias. Não importa saber onde *cada* galáxia elíptica fica no gráfico; a informação pertinente é onde *as* se situam as galáxias elípticas. Para uma categoria, basta saber a posição central e uma medida do seu espalhamento.

Usualmente, se calcula a média para definir o centro e o desvio padrão como dispersão (Figura A.3, à esquerda). No entanto, a média é uma medida que é significativamente afetada pelos pontos divergentes. Para superar isso, dá-se preferência à mediana, que é um estimador muito mais robusto da tendência central de uma amostra. Da mesma maneira, o desvio (da mediana) absoluto médio é uma medida mais robusta da dispersão estatística. Aplicado sobre os nossos dados (Figura A.3, à direita), a correlação entre $u-i$ e γ se torna muito mais regular ao longo da sequência de Hubble.

O método ainda pode ser melhorado. Queremos saber mais precisamente onde e como um tipo morfológico se espalha no espaço das variáveis. Dentro de cada tipo morfológico, a relação entre duas variáveis pode mostrar um comportamento diferente. No modelo precedente, as dispersões foram calculadas diretamente sobre as variáveis e, portanto, as barras de dispersão nos gráficos são horizontais e verticais. Possivelmente, uma direção poderia ser privilegiada, indicando uma correlação para apenas um dado tipo.

A. REPRESENTAÇÕES GRÁFICAS E PONTOS DIVERGENTES

Agora, vamos medir, em cada tipo morfológico, se as variáveis mostram correlações umas com as outras. Essa tarefa é executada pela análise de componentes principais (*Principal components analysis* – PCA). As componentes principais (PCs) são novas variáveis não correlacionadas, combinações lineares das variáveis originais que são parcialmente correlacionadas. As PCs são ortogonais entre si e a primeira PC é a variável com a dispersão máxima. Em duas dimensões, as PCs são os eixos da elipse que melhor se ajusta à nuvem de pontos.

A correlação entre duas variáveis aleatórias $\vec{V} = (X, Y)$ é dada pela matriz de covariância, definida como

$$\text{Cov}(\vec{V}) = E \left[(\vec{V} - E[\vec{V}]) (\vec{V} - E[\vec{V}])^T \right], \quad (\text{A.1})$$

onde $E[\vec{V}]$ é o valor esperado.

Para variáveis aleatórias equiprováveis, $E[\vec{V}]$ é simplesmente a média, e a matriz de covariância se torna

$$\text{Cov}(X, Y) = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}. \quad (\text{A.2})$$

onde σ_X é o desvio padrão de X e σ_X^2 é a variância de X . $\sigma_X\sigma_Y$ é chamada de covariância de X e Y e mede a interdependência das duas variáveis aleatórias.

A matriz de covariância é uma matriz real definida positiva, o que quer dizer que é diagonalizável. Quando a matriz de covariância é diagonal, a covariância entre as variáveis é nula, o que indica que nesse espaço, as variáveis são independentes: estas são as componentes principais. Para calcular as componentes principais basta, portanto, multiplicar as variáveis originais pela matriz de transformação que diagonaliza a matriz de covariância.

Embora seja este o único modo matematicamente correto de se obter as componentes principais, ele tem o inconveniente de não ser estatisticamente robusto. Assim, damos preferência a uma aproximação dos componentes principais que depende menos dos pontos divergentes. Para este fim, definimos uma matriz de pseudo-covariância $\overline{\text{Cov}}$, onde a média é substituída pela mediana:

$$\overline{\text{Cov}}(\vec{V}) = \text{Mediana} \left[(\vec{V} - \text{Mediana}[\vec{V}]) (\vec{V} - \text{Mediana}[\vec{V}])^T \right]. \quad (\text{A.3})$$

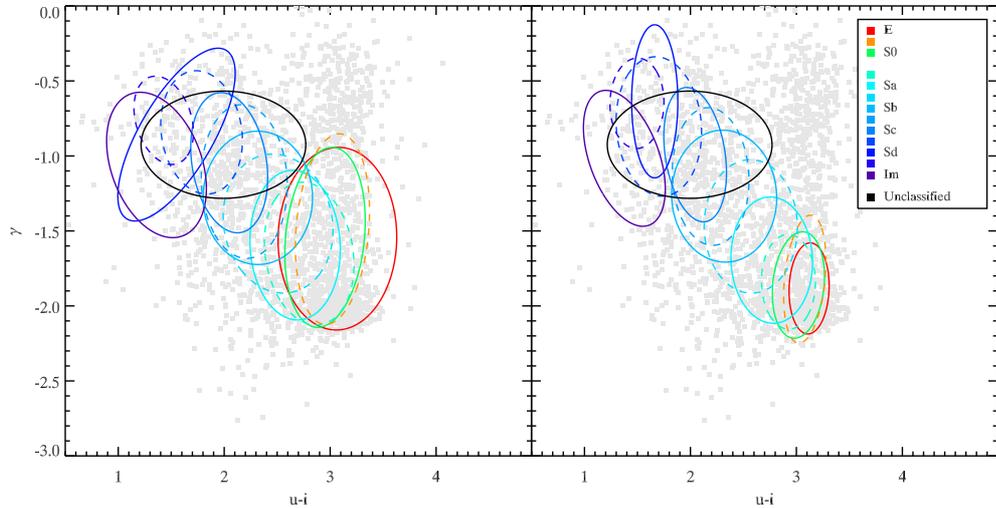


Figura A.4: Plano cor-granulação para os tipos morfológicos. À esquerda: a localização e a dispersão dos tipos morfológicos são representados por elipses cujo eixos são as PCs, com 1σ de comprimento. À direita: os PCs são calculados com a pseudo-covariança, ou seja, com medianas e desvios absolutos medianos.

Aqui, o desvio padrão se torna uma nova grandeza $\bar{\sigma}$, que é proporcional ao desvio (da mediana) absoluto mediano. Para ser coerente com a medida precedente, $\bar{\sigma}$ é corrigido pelo fator 1.4826, que é razão entre o desvio padrão e o desvio absoluto mediano para uma distribuição gaussiana:

$$\bar{\sigma}_X = 1.4826 \cdot \text{Mediana } |X - \text{Mediana}[X]|. \quad (\text{A.4})$$

Na Figura A.4, os dois métodos de cálculo de PCs são comparados. As elipses de cada cor indicam a área do gráfico onde se concentram as galáxias de um mesmo tipo morfológico. Os eixos das elipses são os PCs com 1σ ou $1\bar{\sigma}$. Claramente, as elipses das galáxias *early type* são muito mais expandidas no painel esquerdo.

Podemos finalmente utilizar a PCA para descartar os pontos divergentes. Dentro de cada tipo morfológico, as galáxias fora da elipse com $3\bar{\sigma}$ de semi-eixos estão rejeitadas; usamos os PCs calculados com a matriz de pseudo-covariança. O processo é iterado até que nenhuma galáxia mais seja rejeitada. A amostra fica limpa e, sem os pontos discrepantes, a análise estatística torna-se mais fácil (Figura A.5).

A. REPRESENTAÇÕES GRÁFICAS E PONTOS DIVERGENTES

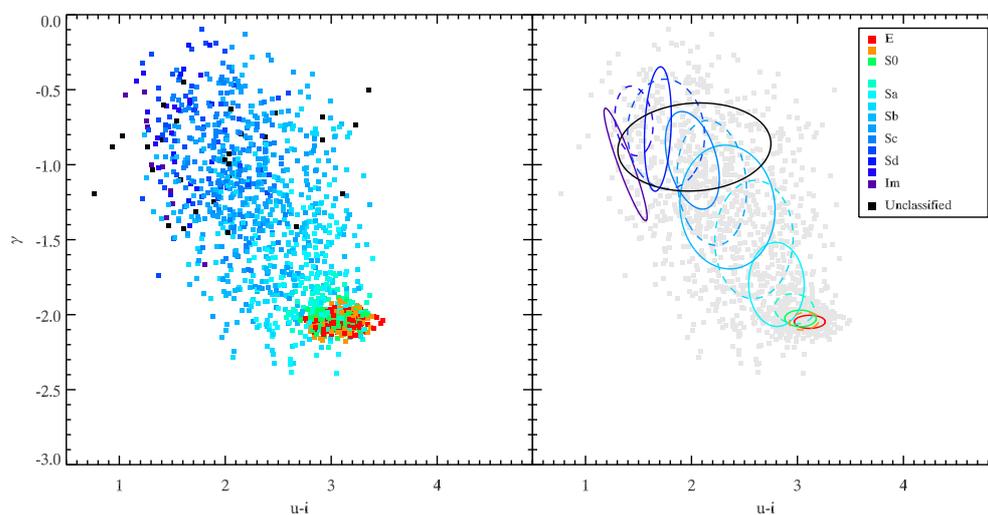


Figura A.5: O plano cor-granulação expurgado dos pontos divergentes a partir da utilização do método PCA, usando pseudo-covariância (esquerdo). Sem pontos divergentes, as elipses podem ser calculadas de maneira normal, com médias e desvios padrões (direita).

Apêndice B

Medindo as desigualdades

O coeficiente de Gini é usado como índice morfológico (Seção 2.4.5). Originalmente, é uma grandeza estatística empregada em economia para medir as desigualdades de riqueza ou de renda em um país. A definição foi adaptada por Abraham et al. (2003) a fim de avaliar as desigualdades entre as intensidades dos pixels de uma galáxia.

Para calcular o coeficiente de Gini, precisa-se conhecer a curva de Lorenz (Lorenz 1905). Em socioeconomia, a curva de Lorenz é uma representação gráfica da distribuição da riqueza total em uma população. Do ponto de vista estatístico, a curva de Lorenz de uma variável aleatória positiva X é a primitiva da inversa da função de distribuição acumulada:

$$L(p) = \frac{1}{\bar{X}} \int_0^p F^{-1}(y)dy, \quad (\text{B.1})$$

onde \bar{X} é a média da variável aleatória, F é a sua função de distribuição acumulada e p é a fração da população mais pobre.

A curva de Lorenz indica qual é a fração da riqueza pertence a uma certa fração da população, ordenada do indivíduo mais pobre ao mais rico.

Para seguir o raciocínio, a curva de Lorenz do Brasil em 2002 está representada à esquerda da Figura B.1. Sobre ela, podemos ler que a metade mais pobre da população fica com apenas 13.74% da renda total; ou que apenas 1% da renda é compartilhada entre os 11% mais pobres da população; ou, ainda, que o 1% mais rico da população ganha 13.14% de toda a renda (dados do IBGE).

Geometricamente, o coeficiente de Gini é o dobro da área entre a diagonal e a curva de Lorenz. A diagonal $y = x$ corresponde à linha de perfeita igualdade, ou seja, a curva

B. MEDINDO AS DESIGUALDADES

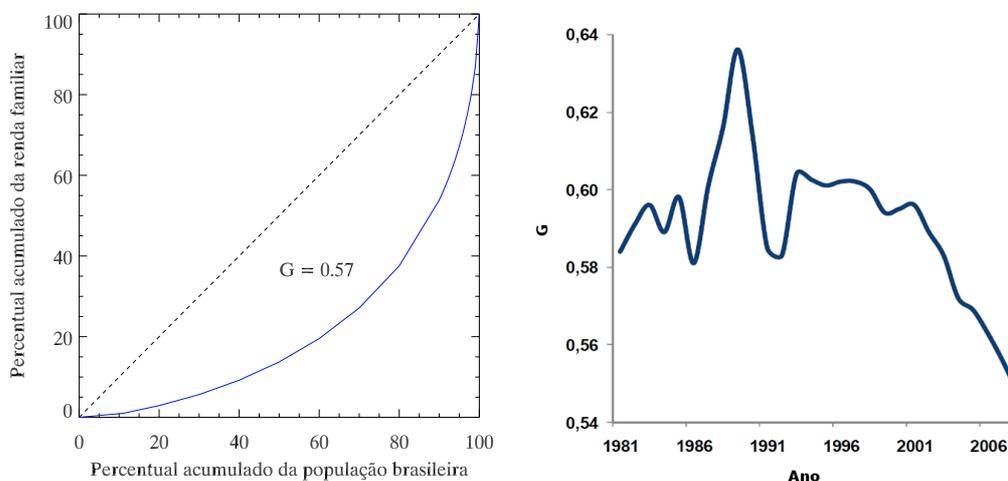


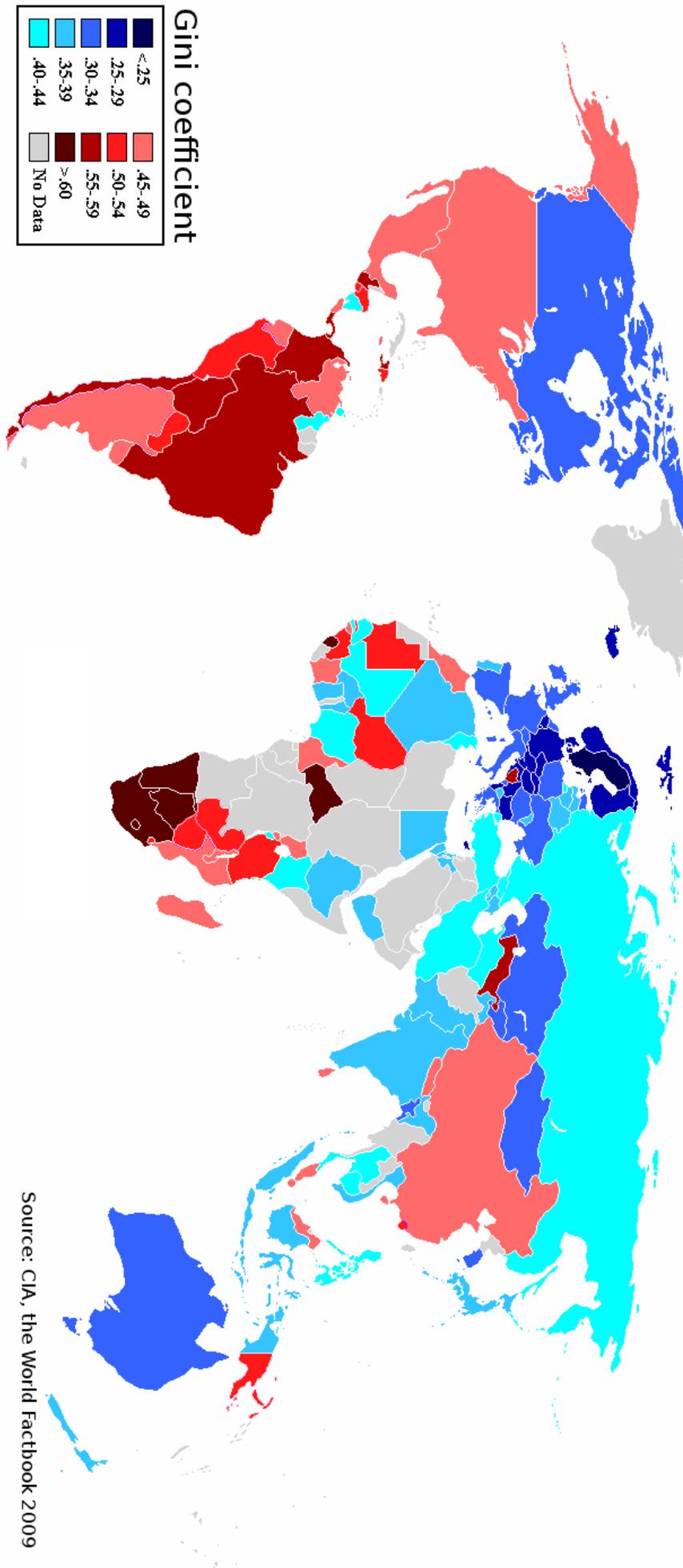
Figura B.1: Esquerda: Curva de Lorenz de renda familiar no Brasil, em 2002 (Fonte: Pesquisa de orçamentos familiares, IBGE). Direita: Evolução do coeficiente de Gini do Brasil durante os últimos anos (Fonte: Pesquisa nacional por amostra de domicílios, IBGE).

de Lorenz gerada por uma variável seguindo uma distribuição uniforme. Quanto mais a curva de Lorenz se afasta da diagonal, com mais desigualdade se distribui a grandeza considerada.

O coeficiente de Gini é uma grandeza que vai de $G=0$, para uma população completamente igual, em que cada indivíduo tem a mesma renda, até $G=1$, para uma população na qual um só indivíduo tem renda enquanto todos os outros não recebem nada¹. A curva de Lorenz representada na Figura B.1 produz um coeficiente de Gini $G=0.5$.

Esse número em si é dificilmente interpretável, mas a sua variação no tempo e no espaço, sim. Assim, a evolução do coeficiente de Gini do Brasil entre 1981 e 2008 revela os progressos sociais realizados nas últimas décadas (Figura B.1, à direita). Mas, em comparação com o resto do mundo, as desigualdades no Brasil ainda são muito elevadas, como o mostra o mapa da Figura B.2.

¹Com muito humor e relevância, Abraham et al. (2003) ilustram esses dois casos extremos: “*In other words, $G=0$ for a perfect communist society and $G=1$ for an absolute monarchy where all riches belong to the king!*”



Source: CIA, the World Factbook 2009

Figura B.2: Mapa do mundo com a cor dos país indicando o coeficiente de Gini.

B. MEDINDO AS DESIGUALDADES

O conceito de desigualdade pode ser generalizado para qualquer variável aleatória. Quando a função de distribuição é desconhecida, a equação B.1 é dificilmente utilizável. Nesse caso, a curva de Lorenz deve ser determinada a partir de todos os eventos (cada valor da variável aleatória), o que envolve os seguintes passos: ordenação, soma e normalização da variável aleatória. Como exemplo, consideramos, em uma população de dez pessoas, duas variáveis: a altura e o rendimento. Valores fictícios são indicados na Tabela B.1.

Tabela B.1: Exemplo de 10 indivíduos cujas alturas e salários são conhecidos.

Indivíduo	A	B	C	D	E	F	G	H	I	J
Altura	170	154	161	145	179	139	171	192	164	175
Rendimento	5200	661	1800	1200	15000	0	639	3100	900	1500

Calculando o coeficiente de Gini, vamos mostrar que a altura é significativamente semelhante para cada um dos dez indivíduos, ao contrário do rendimento. Primeiramente, as sequências das variáveis devem estar ordenadas de maneira crescente. Em seguida, se considera a variável acumulada; a normalização se obtém dividindo pelo total da variável considerada. Essas etapas estão executadas na Tabela B.2.

A curva de Lorenz é a representação da fração da grandeza acumulada em função da fração da população. Assim, na Figura B.3, as curvas de Lorenz são obtidas reproduzindo os valores da quarta linha (pontos azuis) e da oitava linha (pontos vermelhos) em função da última linha da Tabela B.2. As áreas entre a diagonal e a curva de Lorenz valem a metade dos coeficientes de Gini. Temos: $G_{Altura}=0.05$ e $G_{Rendimento}=0.61$. O primeiro coeficiente de Gini é quase nulo, porque a altura de cada um é muito próxima. Mas o segundo mostra disparidades importantes nos rendimentos.

No caso de uma variável discreta, o coeficiente de Gini é a média das diferenças entre cada combinação da variável aleatória (Gini 1912). Devido ao grande número de operações, esse método de cálculo é muito ineficiente em termos de tempo computacional. Assim, prefere-se uma aproximação, conforme Glasser (Glasser 1962), sem referência direta à curva de Lorenz. A maioria dos autores (Abraham et al. 2003; Lisker et al. 2006; Lotz et al. 2004) calcula o coeficiente de Gini com a fórmula de Glasser. Programando os dois métodos, reparamos que o tempo computacional ficava

Tabela B.2: Somas acumuladas e normalizadas das sequências ordenadas.

Sequência ordenada pela altura:										
Indivíduo	F	D	B	C	I	A	G	E	J	H
Altura	139	145	154	161	164	170	171	175	179	192
... acumulada	139	284	438	599	763	933	1104	1279	1458	1650
... e normalizada	0.08	0.17	0.27	0.36	0.46	0.57	0.67	0.78	0.88	1.00
Sequência ordenada pelo peso:										
Indivíduo	F	G	B	I	D	J	C	H	A	E
Rendimento	0	639	661	900	1200	1500	1800	3100	5200	15000
... acumulado	0	639	1300	2200	3400	4900	6700	9800	15000	30000
... e normalizado	0	0.02	0.04	0.07	0.11	0.16	0.22	0.33	0.50	1.00
Índice	1	2	3	4	5	6	7	8	9	10
... normalizado	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

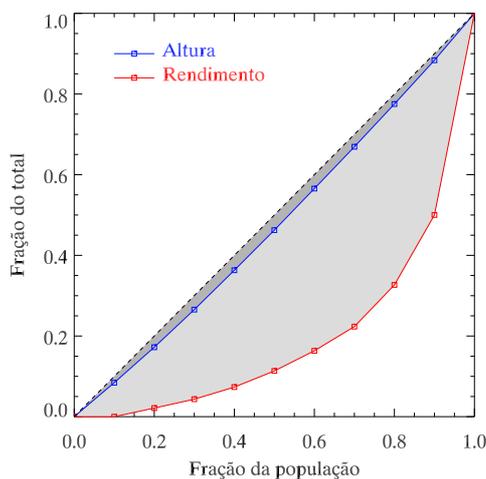


Figura B.3: Curvas de Lorenz para os exemplos de altura (linha azul) e de rendimento (linha vermelha). A linha tracejada é a linha de perfeita igualdade. As áreas cinzas definem os coeficientes de Gini.

B. MEDINDO AS DESIGUALDADES

um pouco menor quando calculávamos explicitamente a curva de Lorenz, como descrito nos parágrafos precedentes. Portanto, não usamos a fórmula de Glasser.

Para aplicar o cálculo do coeficiente de Gini aos pixels das galáxias, algumas precauções têm que ser observadas. A curva de Lorenz – e, conseqüentemente, o coeficiente de Gini – é habitualmente definida somente para variáveis aleatórias estritamente positivas. Se a variável aleatória toma valores negativos, a curva de Lorenz não é mais crescente e o coeficiente de Gini pode atingir valores superiores a 1. Porém, após a subtração do nível do céu, as imagens contêm pixels negativos.

Abraham et al. (2003) propõem pegar o valor absoluto da intensidade dos pixels para calcular o coeficiente de Gini. Lotz et al. (2004) argumentam que essa correção é uma forma de corrigir as desigualdades artificialmente induzidas pelo ruído na imagem.

Estudando a estabilidade do coeficiente de Gini como índice morfológico, Lisker (2008) notou que a razão sinal/ruído domina a medida do coeficiente de Gini quando a abertura é grande demais. Ele recomenda assim a escolha de uma abertura menor, como $2/3 r_{Petro}$. Computamos, então, o coeficiente de Gini dentro de uma máscara formada pela isofota de Petrosian (Seção 2.3.5), que foi reduzida de um terço em cada das duas dimensões. Para todos os outros índices morfológicos, usamos a isofota de Petrosian.

Embora os problemas da abertura e dos pixels negativos sejam resolvidos, o coeficiente de Gini é, sob certas condições, uma medida delicada, que pode ser difícil de se interpretar. Na seção 3.2.5, mostramos que o coeficiente de Gini se correlaciona com o parâmetro de Sérsic e o índice de concentração, e que pode ser considerado como um indicador de tipo concentração. Entretanto, existem exemplos de galáxias com coeficientes de Gini semelhantes, mas com índices de concentração muito diferentes. Na Figura B.4, a imagem de esquerda é uma imagem da galáxia NGC 2175. As duas outras são construídas com os mesmos pixels, de tal maneira que o coeficiente de Gini fica quase inalterado. O índice de concentração, por sua vez, é decrescente da esquerda para direita.

Mesmo com estas incertezas, o coeficiente de Gini é um índice morfológico eficiente. Agora, a dispersão estatística, tal como medida, por exemplo, pelo desvio padrão, também é uma indicação de desigualdade. A dispersão estatística, todavia, é independente da média da distribuição. Na Tabela B.1, dá para prever, sem calcular o

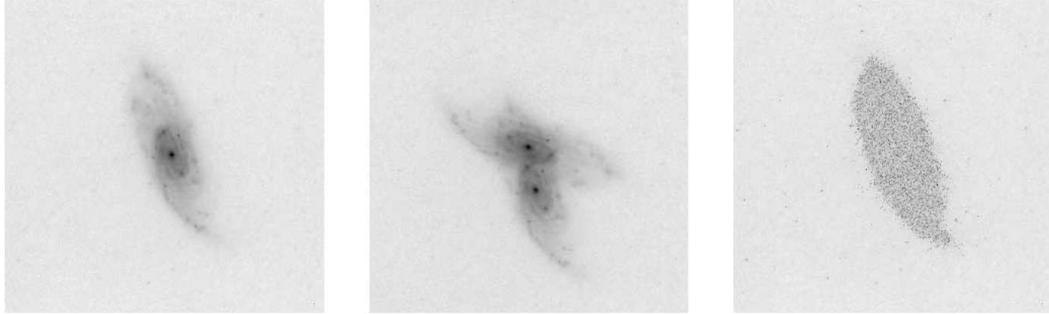


Figura B.4: Três “galáxias” com coeficientes de Gini basicamente idênticos. Esquerda: a galáxia NGC 2715. Meio: Imagem obtida por justaposição da galáxia original com uma mesma galáxia ligeiramente virada e deslocada. Direita: Imagem construída permutando aleatoriamente os pixels da galáxia. Figuras tiradas de Abraham et al. (2003).

coeficiente de Gini, que as desigualdades são muito menores para a altura que para o rendimento, porque todos os indivíduos têm uma altura ao redor da altura média, o que não é o caso do rendimento. Mas se o rendimento de cada um fosse aumentado de um mesmo montante, bastante grande em relação ao rendimento inicial, as desigualdades diminuiriam, sem que a dispersão fosse alterada. Isso ilustra bem a importância da média, que desempenha o papel de ponto de referência da distribuição.

Fundamentalmente, a noção de desigualdade se refere à ideia de espalhamento relativo. Podemos, então, usar como indicador de desigualdade o espalhamento relativo das intensidades, ou seja, a dispersão normalizada pela média: σ_I/\bar{I} .

Aqui, σ_I é escolhido como o desvio absoluto médio, em vez do desvio padrão:

$$\sigma_I = \frac{1}{N-1} \sum_{i=1}^N |I_i - \bar{I}|. \quad (\text{B.2})$$

Usamos o desvio absoluto médio, e não o desvio padrão, porque reparamos uma melhor correlação com o coeficiente de Gini. A Figura B.5 mostra a grandeza σ_I/\bar{I} em função do coeficiente de Gini calculada para as galáxias do catálogo morfológico de Fukugita et al. (2007). A correlação é tão forte, sem nenhum ponto divergente, que essa grandeza poderia ser usada como alternativa simples e rápida ao coeficiente de Gini como medida das desigualdades e, portanto, como índice morfológico.

Assim, historicamente, a quantificação da concentração radial da luz nas galáxias percorreu um longo caminho. Estudando as “nebulosas elípticas”, de Vaucouleurs

B. MEDINDO AS DESIGUALDADES

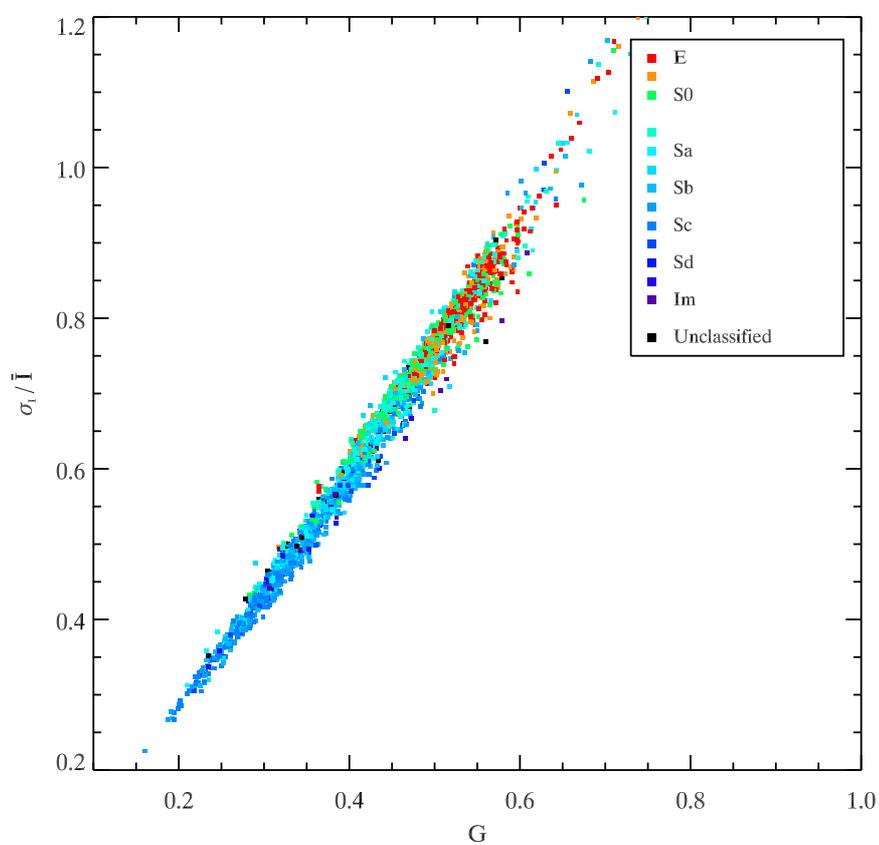


Figura B.5: O coeficiente de Gini G é muito bem aproximado pelo quociente σ_1 / \bar{I} .

(1948) ajustou uma função empírica aos perfis radiais; Sérsic (1963) generalizou a lei de de Vaucouleurs, colocando um parâmetro adicional; para não assumir um modelo subjacente, Kent (1985) definiu um indicador não paramétrico através do índice de concentração; com o propósito de superar a necessidade de definir um centro galático, Abraham et al. (2003) usam um indicador das desigualdades sociais – o coeficiente de Gini –, em vez do índice de concentração; finalmente, mostramos que a grandeza σ_I/\bar{I} equivale ao coeficiente de Gini.

Em nosso estudo morfológico, especialmente na seção 3.2.5, o coeficiente de Gini é calculado a partir da definição, ou seja, com auxílio da função de Lorenz, como apresentado acima.

B. MEDINDO AS DESIGUALDADES

Apêndice C

Wavelet vs Fourier

A transformada de *wavelet*¹ é um método geral de análise de dados que pode ser implementada a uma ou muitas dimensões. Embora diferente em vários pontos, a transformada de Fourier pode fornecer um resultado semelhante à transformada de *wavelet*.

A principal diferença entre as transformadas de *wavelet* e de Fourier é que, na transformada de Fourier, as funções usadas como base do novo espaço – senos e cossenos – são funções periódicas. Isso tem como consequência a perda de localização no espaço de Fourier. O resultado da transformada de Fourier é uma função complexa dependente só da frequência, e cuja a parte real é a amplitude e a parte imaginária é a fase.

Na transformada de *wavelet*, as bases do novo espaço – as próprias *wavelets* – são funções compactas, definidas não nulas apenas sobre um intervalo finito chamado de suporte. As *wavelets* podem ser de várias formas, mas devem ser oscilantes, e a distância entre dois extremos será interpretada como um pseudo-período T , que corresponde a uma pseudo-frequência ν . Essas grandezas são representadas na Figura C.1 sobre dois exemplos de *wavelets*: à esquerda, a *wavelet* associada à função de escala linear; à direita, a *wavelet* associada à função de escala B_3 -spline (Veja apêndice do artigo do Capítulo 4 para maiores detalhes).

Agora, como uma *wavelet* não é nem periódica, nem infinita, não se pode definir uma fase. Em seu lugar, a transformada de *wavelet* conserva a localização da *wavelet*. E, ao contrário da transformada de Fourier, a amplitude é definida em função da loca-

¹A tradução portuguesa de *wavelet* seria ondinha

C. WAVELET VS FOURIER

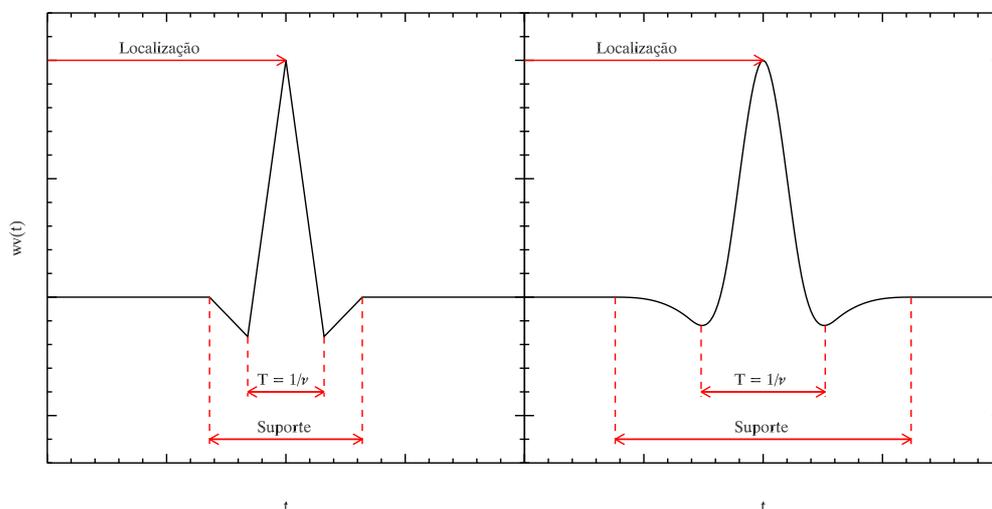


Figura C.1: Dois exemplos de *wavelet*.

lização, e isso para cada frequência. Conseqüentemente, o resultado da transformada de *wavelet* é uma função de duas variáveis: a frequência e a localização.

Chama-se de coeficientes de *wavelet* as amplitudes em função da localização. Há um coeficiente de *wavelet* por frequência. Um coeficiente de *wavelet* é o resultado de um produto de convolução do sinal de entrada com a *wavelet* da frequência considerada.

Quando o sinal de entrada não é temporal, mas espacial, como é o caso para imagens, se fala, às vezes, de frequência espacial. A localização é uma grandeza que tem mesmas unidade e dimensão do que a variável do sinal de entrada, o que simplifica a análise e a interpretação dos coeficientes de *wavelet*. Uma convolução por uma *wavelet* age como um filtro passa-faixa. Nos exemplos da Figura C.1, as frequências maiores do que ν e menores do que $\nu/2$ são removidas. Uma família de *wavelets* é então um banco de filtros que isola em um sinal as frequências desejadas.

Mesmo se a transformada de Fourier perde a informação geográfica, é possível usá-la para cortar as frequências. Para isso, temos que escolher um banco de filtros que será aplicado no espaço de Fourier antes de reconstruir o sinal filtrado com a transformada inversa de Fourier.

O filtro mais simples é a função retangular, que vale 1 para as frequências que queremos conservar e 0 para aquelas que queremos cortar. Para filtros mais suaves,

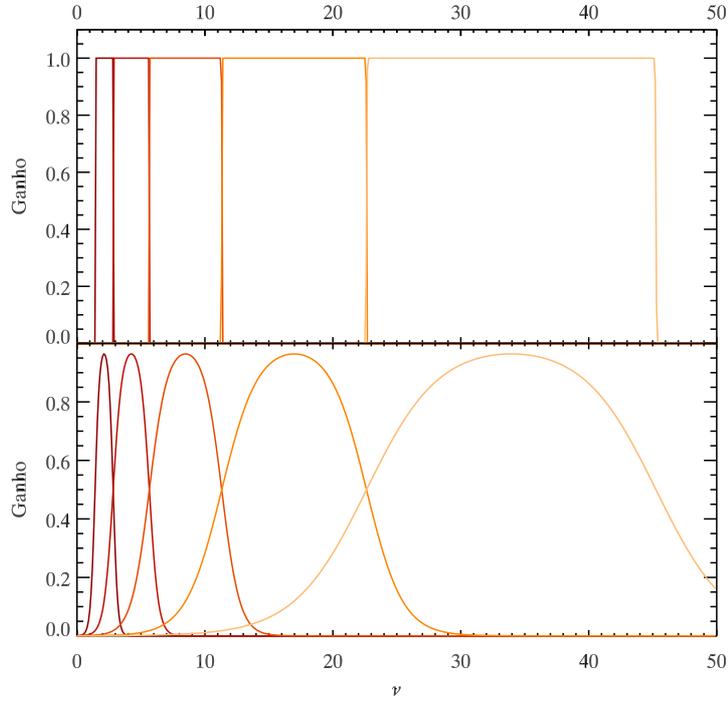


Figura C.2: Dois bancos de filtros. O primeiro é composto de funções retangulares, o segundo de diferenças de duas tangentes hiperbólicas.

podemos, por exemplo, construir uma função como:

$$f(\nu) = \frac{1}{2}(\tanh(k\nu_1) - \tanh(k\nu_2)) \quad (\text{C.1})$$

onde ν_1 e ν_2 são as frequências de corte, e k é um parâmetro que define quão abrupta é a função; $k = \infty$ leva à função retangular. Esses filtros com $k = 4$ são plotados na Figura C.2, assim como os filtros retangulares.

A imagem *Fukugita1130* é decomposta conforme os três métodos descritos acima, e os resultados são mostrados na Figura C.3. Na primeira coluna, aplica-se à imagem o algoritmo da transformada rápida de Fourier (*Fast Fourier Transform* – FFT); em seguida as frequências são filtradas usando as funções retangulares da Figura C.2; finalmente, as imagens são reconstruídas usando a FFT inversa. Na segunda coluna, as frequências no espaço de Fourier são selecionadas com o segundo banco de filtros da Figura C.2. Na terceira coluna, a decomposição é obtida pela transformada de *wave-*

C. WAVELET VS FOURIER

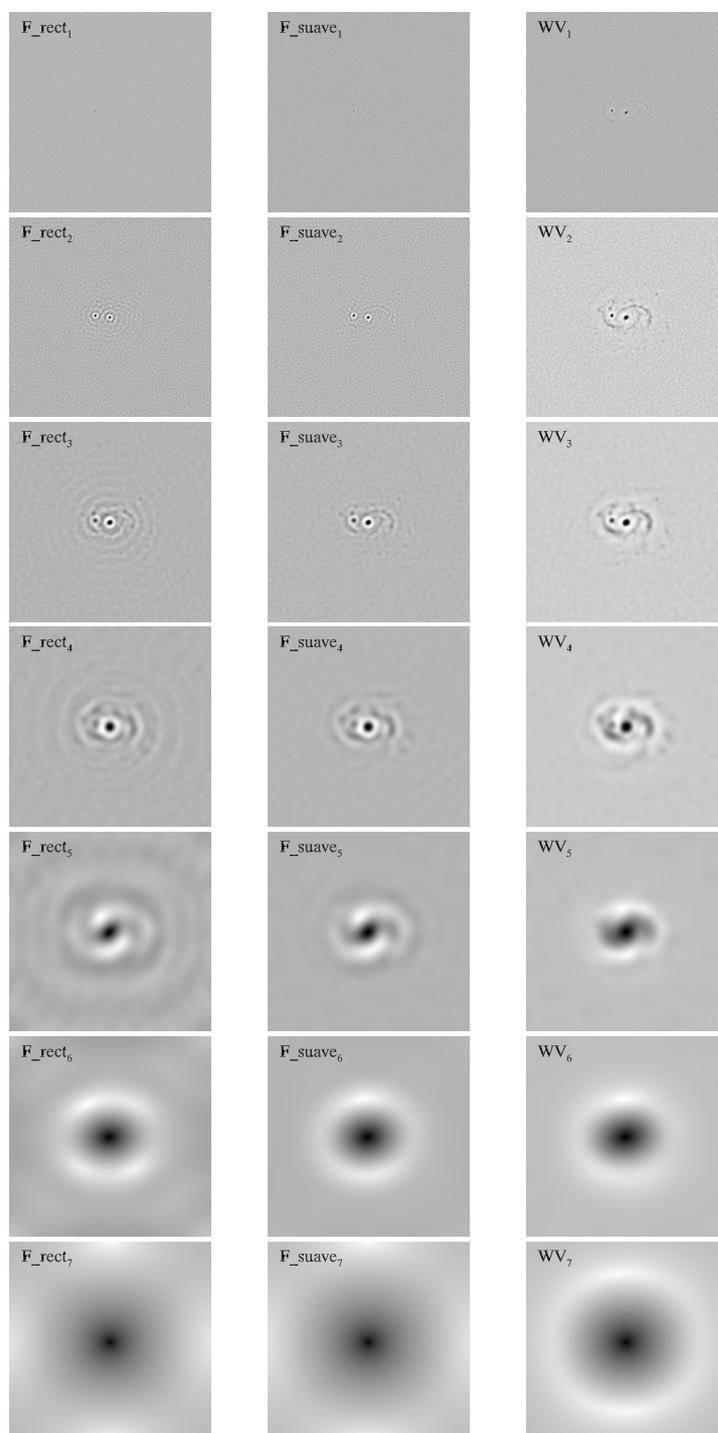


Figura C.3: Três exemplos de decomposição da imagem *Fukugita1130*: com a transformada de Fourier e o banco de filtros retangulares (1^a coluna), com a transformada de Fourier e o banco de filtros tangentes hiperbólicas (2^a coluna), e com a transformada de *wavelet* (3^a coluna).

let, usando o algoritmo “à trous” e a função de escala B_3 -spline, ou seja, com *wavelets* semelhantes àquela no gráfico direito da Figura C.1.

As três decomposições são muitas parecidas. A descontinuidade dos filtros retangulares cria artefatos ao redor dos picos luminosos, um pouco como ondas circulares na superfície da água. Este problema desaparece com os filtros suaves. Mas o melhor resultado é produzido pela transformada de *wavelet*: oferece mais detalhes nas altas frequências espaciais, ou seja, nas pequenas escalas. Além disso, a transformada de *wavelet* é, em média, duas vezes mais rápido.

Em nosso trabalho, a transformada de *wavelets* fica na origem do índice de granulação (seção). Os coeficientes de *wavelets*

C. WAVELET VS FOURIER

Referências

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *The Astrophysical Journal Supplement Series*, 182, 543 [14](#)
- Abraham, R. G., van den Bergh, S., & Nair, P. 2003, *ApJ*, 588, 218 [xviii](#), [48](#), [62](#), [147](#), [148](#), [150](#), [152](#), [153](#), [155](#)
- Aihara, H., Allende Prieto, C., An, D., et al. 2011, *ApJS*, 193, 29 [14](#)
- Aller, L. H., ed. 1984, *Astrophysics and Space Science Library*, Vol. 112, *Physics of thermal gaseous nebulae* [18](#)
- Andrae, R., Jahnke, K., & Melchior, P. 2011a, *MNRAS*, 411, 385 [84](#)
- Andrae, R., Melchior, P., & Jahnke, K. 2011b, *ArXiv e-prints* [52](#)
- Arp, H. 1966, *ApJS*, 14, 1 [6](#)
- Baillard, A., Bertin, E., Mellier, Y., et al. 2006, in *Astronomical Society of the Pacific Conference Series*, Vol. 351, *Astronomical Data Analysis Software and Systems XV*, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, 236–+ [19](#)
- Benson, A. J. 2010, *Phys. Rep.*, 495, 33 [7](#)
- Bershady, M. A., Jangren, A., & Conselice, C. J. 2000, *AJ*, 119, 2645 [48](#)
- Bertin, E. & Arnouts, S. 1996, *A&AS*, 117, 393 [19](#), [40](#)
- Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. 2003, *ApJ*, 594, 186 [52](#)
- Buta, R. & Combes, F. 1996, *Fund. Cosmic Phys.*, 17, 95

REFERÊNCIAS

- Buta, R. J. 2011, ArXiv e-prints [7](#)
- Conselice, C. J. 2003, ApJS, 147, 1 [48](#), [84](#), [104](#)
- Conselice, C. J., Bershad, M. A., Dickinson, M., & Papovich, C. 2003, AJ, 126, 1183
[84](#)
- de Vaucouleurs, G. 1948, Annales d'Astrophysique, 11, 247 [49](#), [153](#)
- de Vaucouleurs, G. 1958, ApJ, 128, 465 [49](#)
- de Vaucouleurs, G. 1959, Handbuch der Physik, 53, 275 [xiii](#), [5](#)
- de Vaucouleurs, G., de Vaucouleurs, A., Corwin, Jr., H. G., et al. 1991, Third Reference Catalogue of Bright Galaxies, ed. de Vaucouleurs, G., de Vaucouleurs, A., Corwin, H. G., Jr., Buta, R. J., Paturel, G., & Fouque, P. [6](#), [16](#), [85](#)
- de Vaucouleurs, G., de Vaucouleurs, A., & Corwin, J. R. 1976, in Second reference catalogue of bright galaxies, 1976, Austin: University of Texas Press., 0 [6](#)
- de Vaucouleurs, G. H., de Vaucouleurs, A., & Shapley, H. 1964, Reference catalogue of bright galaxies, ed. de Vaucouleurs, G. H., de Vaucouleurs, A., & Shapley, H. [xiii](#), [5](#), [6](#), [16](#)
- Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72 [14](#)
- Fortson, L., Masters, K., Nichol, R., et al. 2011, ArXiv e-prints [45](#)
- Fukugita, M., Nakamura, O., Okamura, S., et al. 2007, AJ, 134, 579 [xv](#), [13](#), [15](#), [17](#), [20](#), [23](#), [46](#), [55](#), [85](#), [141](#), [153](#)
- Gini, C. 1912, Variabilità e mutabilità, ed. Gini, C. (Libreria Eredi Virgilio Veschi, Roma) [150](#)
- Glasser, G. J. 1962, Journal of the American Statistical Association, 57, 648 [150](#)
- Graham, A. W. & Driver, S. P. 2005, PASA, 22, 118 [49](#)
- Graham, A. W. & Driver, S. P. 2007, ApJ, 655, 77 [52](#)

- Hodge, P. W. 1966, *The physics and astronomy of galaxies and cosmology*, ed. Hodge, P. W. 5
- Holwerda, B. W. 2005, *ArXiv Astrophysics e-prints* 38
- Hubble, E. P. 1926, *ApJ*, 64, 321 4, 45
- Hubble, E. P. 1936, *Realm of the Nebulae*, ed. Hubble, E. P. 4
- Huertas-Company, M., Rouan, D., Tasca, L., Soucail, G., & Le Fèvre, O. 2008, *A&A*, 478, 971 84
- Izotov, Y. I., Stasińska, G., Meynet, G., Guseva, N. G., & Thuan, T. X. 2006, *A&A*, 448, 955 13, 18, 23
- Kennicutt, Jr., R. C. 1989, *ApJ*, 344, 685 9
- Kent, S. M. 1985, *ApJS*, 59, 115 48, 155
- Kniazev, A. Y., Pustilnik, S. A., Grebel, E. K., Lee, H., & Pramskij, A. G. 2004, *ApJS*, 153, 429 13, 18, 23
- Kron, R. G. 1980, *ApJS*, 43, 305 40
- Lintott, C., Schawinski, K., Bamford, S., et al. 2011, *MNRAS*, 410, 166 46
- Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, *MNRAS*, 389, 1179 45
- Lisker, T. 2008, *ApJS*, 179, 319 152
- Lisker, T., Debattista, V. P., Ferreras, I., & Erwin, P. 2006, *MNRAS*, 370, 477 150
- Lorenz, M. O. 1905, *Publications of the American Statistical Association*, Volume 9, Number 70, p. 209-219, 9, 209 147
- Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2008, *MNRAS*, 391, 1137 9
- Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2010a, *MNRAS*, 404, 590 9
- Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2010b, *MNRAS*, 404, 575 9

REFERÊNCIAS

- Lotz, J. M., Primack, J., & Madau, P. 2004, *AJ*, 128, 163 [48](#), [69](#), [84](#), [104](#), [150](#), [152](#)
- Lupton, R., Gunn, J. E., Ivezić, Z., Knapp, G. R., & Kent, S. 2001, in *Astronomical Society of the Pacific Conference Series*, Vol. 238, *Astronomical Data Analysis Software and Systems X*, ed. F. R. Harnden Jr., F. A. Primini, & H. E. Payne, 269–+ [19](#)
- Morgan, W. W. 1958, *PASP*, 70, 364 [5](#)
- Nair, P. B. & Abraham, R. G. 2010, *ApJS*, 186, 427 [xvi](#), [7](#), [85](#), [86](#)
- Ngan, W., van Waerbeke, L., Mahdavi, A., Heymans, C., & Hoekstra, H. 2009, *MNRAS*, 396, 1211 [52](#)
- Nilson, P. 1973, *Uppsala general catalogue of galaxies*, ed. Nilson, P. [6](#)
- Pagel, B. E. J., Edmunds, M. G., Blackwell, D. E., Chun, M. S., & Smith, G. 1979, *MNRAS*, 189, 95 [18](#)
- Park, C. & Choi, Y.-Y. 2005, *ApJ*, 635, L29 [84](#)
- Parodi, B. R. & Binggeli, B. 2003, *A&A*, 398, 501 [59](#)
- Petrosian, V. 1976, *ApJ*, 209, L1 [24](#), [40](#)
- Pilyugin, L. S. & Thuan, T. X. 2005, *ApJ*, 631, 231 [18](#), [125](#)
- Refregier, A. 2003, *MNRAS*, 338, 35 [52](#)
- Sandage, A. 1961, *The Hubble atlas of galaxies*, ed. Sandage, A. [16](#)
- Sandage, A. 1975, *Classification and Stellar Content of Galaxies Obtained from Direct Photography*, ed. Sandage, A., Sandage, M., & Kristian, J. (the University of Chicago Press), 1–+ [2](#)
- Sandage, A. & Brucato, R. 1979, *AJ*, 84, 472 [17](#)
- Sandage, A. & Tammann, G. A. 1980, *A revised Shapley-Ames Catalog of bright galaxies*, ed. Sandage, A. & Tammann, G. A. [6](#)

- Scarlata, C., Carollo, C. M., Lilly, S., et al. 2007, *ApJS*, 172, 406 [84](#)
- Schade, D., Lilly, S. J., Crampton, D., et al. 1995, *ApJ*, 451, L1+ [48](#)
- Sérsic, J. L. 1963, *Boletín de la Asociación Argentina de Astronomía La Plata Argentina*, 6, 41 [48](#), [49](#), [155](#)
- Shamir, L. 2009, *MNRAS*, 399, 1367 [84](#)
- Storrie-Lombardi, M. C., Lahav, O., Sodre, Jr., L., & Storrie-Lombardi, L. J. 1992, *MNRAS*, 259, 8P [84](#)
- Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, *AJ*, 123, 485 [13](#), [19](#)
- Takamiya, M. 1999, *ApJS*, 122, 109 [48](#)
- Tody, D. 1993, in *Astronomical Society of the Pacific Conference Series*, Vol. 52, *Astronomical Data Analysis Software and Systems II*, ed. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes, 173–+ [19](#)
- Toomre, A. & Toomre, J. 1972, *ApJ*, 178, 623 [xiii](#), [9](#)
- van den Bergh, S. 1960a, *ApJ*, 131, 558 [5](#)
- van den Bergh, S. 1960b, *ApJ*, 131, 215 [5](#)
- von Goethe, J. W. 1817, *Zur Naturwissenschaft {überhaupt, besonders zur Morphologie}*.
No. vol. 1 (J. G. Cotta, Stuttgart)
- Vorontsov-Velyaminov, B. A. 1959, in *Atlas and catalog of interacting galaxies (1959)*,
[0](#) [6](#)
- Westera, P., Cuisinier, F., Curty, D., & Buser, R. 2012, *MNRAS*, 2372 [13](#), [18](#)
- Wolf, M. 1908, *Publikationen des Astrophysikalischen Instituts Koenigstuhl-Heidelberg*, 3, 109 [xiii](#), [2](#), [3](#)
- York, D. G., Adelman, J., Anderson, J. J. E., et al. 2000, *AJ*, 120, 1579 [13](#), [19](#)
- Zamojski, M. A., Schiminovich, D., Rich, R. M., et al. 2007, *ApJS*, 172, 468 [52](#)