

!ojo! Tomografia PCA aplicado às galáxias do Projeto CALIFA Survey

Eduardo Alberto Duarte Lacerda

Orientador:

Prof. Dr. Roberto Cid Fernandes Jr.

• • •

Universidade Federal de Santa Catarina
Centro de Ciências Físicas e Matemáticas
Curso de Pós-Graduação em Física

Dissertação de mestrado apresentada ao Curso de
Pós-Graduação em Física da UFSC em preenchimen-
to parcial dos requisitos para obtenção do tí-
tulo de Mestre em Física.

Florianópolis (SC) – 24 de fevereiro de 2014

Sumário

1	Introdução	1
1.1	Um labirinto de dados	1
1.2	A nova ferramenta - Tomografia PCA	5
1.3	Este trabalho	6
1.3.1	Organização deste trabalho	7
2	O projeto CALIFA e o <i>pipeline</i> PyCASSO	8
2.1	O survey CALIFA	8
2.2	As <i>pipelines</i> QBICK e PyCASSO	11
2.2.1	QBICK	11
2.2.2	PyCASSO	13
2.2.3	Exemplos de utilização do PyCASSO	15
3	PCA e Tomografia PCA	19
3.1	Principal Component Analisys	19
3.1.1	PCA das galáxias do CALIFA	20
3.2	Tomografia PCA	23
3.3	Exemplos de resultado da Tomografia PCA	23
3.3.1	Descoberta de linhas largas em um LINER	23

3.3.2	Reflexao da luz de um AGN escondido em NGC 7097	24
3.4	Tomografia PCA para dados CALIFA: diferenças com respeito a trabalhos anteriores	26
4	Aplicação a cubos de dados CALIFA	28
4.1	PCA dos espectros observados: Efeitos de amplitude	28
4.2	PCA dos espectros sintéticos	33
4.3	PCA logarítmica	37
4.4	Comparando as PCs com o <i>STARLIGHT</i> : engenharia reversa	38
5	Aplicando a Tomografia PCA em 8 galáxias do CALIFA	45
5.1	Apresentação das galáxias	45
5.2	Galáxias espirais	46
5.2.1	NGC 0001 - CALIFA 8	46
5.2.2	NGC 0776 - CALIFA 73	47
5.2.3	NGC 4210 - CALIFA 518	49
5.3	Galáxias <i>early-type</i>	66
5.3.1	NGC 1167 - CALIFA 119	66
5.3.2	NGC 6515 - CALIFA 864	66
5.4	<i>Mergers</i>	79
5.4.1	NGC 2623 - CALIFA 213	79
5.4.2	ARP 220 - CALIFA 802	79
6	Conclusões e perspectivas	92
6.1	Este trabalho	92
6.2	Trabalhos futuros	92
6.2.1	Linhas de emissão e intervalos específicos em comprimento de onda .	92

A [FIXME] Python	93
-------------------------	-----------

Referências Bibliográficas	I
-----------------------------------	----------

Lista de Figuras

1.1	Espectro médio e 7 primeiras PCs de uma biblioteca de modelos.	4
2.1	Diagrama cor-magnitude para as galáxias do CALIFA.	10
2.2	Configuração do <i>bundle</i> de fibras do PPMAS/PPAK.	10
2.3	Exemplo de máscaras em um espectro do cubo de dados.	12
2.4	!○j○!	13
2.5	Exemplo de programa utilizando PyCASSO	15
2.6	Programa idade estelar média	16
2.7	Mapa da idade estelar média da galáxia NGC 2916	16
2.8	Exemplo de programa para perfil radial	17
2.9	Perfil radial da idade estelar média da galáxia NGC 2916	17
3.1	Exemplo de cálculo de PCA usando o PyCASSO e SciPy	22
3.2	<i>Scree test</i> na galáxia NGC 4736.	24
3.3	Tomograma e autoespectro 1 da galáxia NGC 4736.	25
3.4	Tomograma e autoespectro 2 da galáxia NGC 4736.	25
3.5	Tomograma e autoespectro 3 da galáxia NGC 4736.	26
4.1	!○j○!	29
4.2	Tomogramas de 1 a 5 da galáxia NGC 2916 - F_{obs} .	31

4.3	Tomogramas de 1 a 5 da galáxia NGC 2916 - $F_{obs}/F_{\lambda 5365}$	32
4.4	Fluxos de normalização para cada zona da galáxia K0277.	33
4.5	Scree test comparativo entre 3 PCAs.	34
4.6	Tomogramas de 1 a 5 da galáxia NGC 2916 - $F_{syn}/F_{\lambda 5365}$	36
4.7	Correlações PCs vs. parâmetros físicos - F_{obs} norm.	40
4.8	Correlações PCs vs. parâmetros físicos - F_{syn} norm.	41
4.9	Dados no espaço das PCs vs AV- F_{obs} norm.	42
4.10	Dados no espaço das PCs vs AV- F_{syn} norm.	43
4.11	Correlações PCs vs. parâmetros físicos - F_{obs}	44
5.1	Propriedades físicas da galáxia NGC 0001.	48
5.2	Scree test comparativo entre 3 PCAs - NGC 0001.	49
5.3	Tomogramas de 1 a 5 para o cubo F_{obs} norm. - NGC 0001.	50
5.4	Tomogramas de 1 a 5 para o cubo F_{syn} norm. - NGC 0001.	51
5.5	Correlações PCs vs. parâmetros físicos - F_{obs} norm. - NGC 0001	52
5.6	Correlações PCs vs. parâmetros físicos - F_{syn} norm. - NGC 0001	53
5.7	Propriedades físicas da galáxia NGC 0776.	54
5.8	Scree test comparativo entre 3 PCAs - NGC 0776.	55
5.9	Tomogramas de 1 a 5 para o cubo F_{obs} norm. - NGC 0776.	56
5.10	Tomogramas de 1 a 5 para o cubo F_{syn} norm. - NGC 0776.	57
5.11	Correlações PCs vs. parâmetros físicos - F_{obs} norm. - NGC 0001	58
5.12	Correlações PCs vs. parâmetros físicos - F_{syn} norm. - NGC 0001	59
5.13	Propriedades físicas da galáxia NGC 4210.	60
5.14	Scree test comparativo entre 3 PCAs - NGC 4210.	61
5.15	Tomogramas de 1 a 5 para o cubo F_{obs} norm. - NGC 4210.	62

5.16 Tomogramas de 1 a 5 para o cubo F_{syn} norm. - NGC 4210.	63
5.17 Correlações PCs vs. parâmetros físicos - F_{obs} norm. - NGC 4210.	64
5.18 Correlações PCs vs. parâmetros físicos -um padrão de velocidades F_{syn} norm. - NGC 4210.	65
5.19 Propriedades físicas da galáxia NGC 1167.	67
5.20 Scree test comparativo entre 3 PCAs - NGC 1167.	68
5.21 Tomogramas de 1 a 5 para o cubo F_{obs} norm. - NGC 1167.	69
5.22 Tomogramas de 1 a 5 para o cubo F_{syn} norm. - NGC 1167.	70
5.23 Correlações PCs vs. parâmetros físicos - F_{obs} norm. - NGC 1167.	71
5.24 Correlações PCs vs. parâmetros físicos - F_{syn} norm. - NGC 1167.	72
5.25 Propriedades físicas da galáxia NGC 6515.	73
5.26 Scree test comparativo entre 3 PCAs - NGC 6515.	74
5.27 Tomogramas de 1 a 5 para o cubo F_{obs} norm. - NGC 6515.	75
5.28 Tomogramas de 1 a 5 para o cubo F_{syn} norm. - NGC 6515.	76
5.29 Correlações PCs vs. parâmetros físicos - F_{obs} norm. - NGC 6515.	77
5.30 Correlações PCs vs. parâmetros físicos - F_{syn} norm. - NGC 6515.	78
5.31 Propriedades físicas da galáxia NGC 2623.	80
5.32 Scree test comparativo entre 3 PCAs - NGC 2623.	81
5.33 Tomogramas de 1 a 5 para o cubo F_{obs} norm. - NGC 2623.	82
5.34 Tomogramas de 1 a 5 para o cubo F_{syn} norm. - NGC 2623.	83
5.35 Correlações PCs vs. parâmetros físicos - F_{obs} norm. - NGC 2623.	84
5.36 Correlações PCs vs. parâmetros físicos - F_{syn} norm. - NGC 2623.	85
5.37 Propriedades físicas da galáxia ARP 220.	86
5.38 Scree test comparativo entre 3 PCAs - ARP 220.	87
5.39 Tomogramas de 1 a 5 para o cubo F_{obs} norm. - ARP 220.	88

5.40 Tomogramas de 1 a 5 para o cubo F_{obs} norm. - ARP 220.	89
5.41 Correlações PCs vs. parâmetros físicos - F_{obs} norm. - ARP 220.	90
5.42 Correlações PCs vs. parâmetros físicos - F_{syn} norm. - ARP 220.	91

Lista de Tabelas

2.1	Relação de pixels e zonas em algumas galáxias do CALIFA	14
5.1	Relação de galáxias do CALIFA usadas neste trabalho.	46

Capítulo 1

Introdução

1.1 Um labirinto de dados

Cientistas hoje, em sua maioria, encontram-se emaranhados em meio a um labirinto de informações. Grandes corredores com bilhões de intersecções interligadas que funcionam como uma fantástica e infinita biblioteca de dados. Um dia formada por livros, hoje formada de 0s e 1s que, através de infinitas combinações, juntos abarcam o combustível para o desenvolvimento do conhecimento da huminadade. Tal qual seja sua área, você tem, já teve, ou terá um dia que recorrer a algum banco de dados de qualquer espécie pois hoje são parte fundamental no futuro da pesquisa científica.

Os primeiros *surveys*¹ astronômicos surgem com a inata curiosidade do homem de observar tudo a sua volta e registrar suas observações. O crescimento desses catálogos é produto direto da evolução dos equipamentos usados nestas investigações. No início da década de 80 Huchra et al. (1983) constroem um *redshift-survey* que coletou espectros para 2500 galaxias dando início a produção sistemática de dados na astronomia extragalática. Esse trabalho também foi seguidos por outros trabalhos (e.g., Huchra 1988; da Costa et al. 1988) e de lá para cá a quantidade (e a qualidade) de dados só aumentou. Com a criação dos *mega-surveys* (SDSS; York et al. 2000) (2dFGRS; Colless 1999) (2MASS; Skrutskie et al. 2006), alguns já terminados, outros ainda começando ou por terminar (LSST; Ivezic et al. 2008) (J-PAS; Benítez et al. 2009), ultrapassou-se o patamar do humanamente impossível de se analisar individualmente cada objeto, necessitando assim a ajuda de máquinas e métodos computacionais cada vez mais eficientes. Com esse crescimento exponencial na quantidade de dados, precisamos cada vez

¹Um *survey* astronômico é um levantamento de informações ou mapeamento de regiões do céu utilizando telescópios e detectores.

mais de ferramentas matemáticas/estatísticas, e essa dissertação trata precisamente disso.

Um tipo de clássico de ferramenta de análise são os programas que ajustam os dados segundo algum modelo, extraíndo dessa análise informações de valor astrofísico. No contexto de espectros de galáxias, tema desta dissertação, uma ferramenta deste estilo é o código STARLIGHT, desenvolvido por Cid Fernandes et al. (2005), que decompõe o espectro observado em termos de populações estelares de distintas idades e metalicidades em um processo conhecido como síntese espectral. A aplicação desse método a quase um milhão de espectros de galáxias do SDSS gerou uma série de resultados **!obj!** (e.g., algumas REFS!!). O STARLIGHT, assim como várias outras ferramentas similares **!obj!** (Panter et al ..., Gallazzi, Ocvirk ...), propõe uma pergunta bem definida ("qual a história de formação de uma dada galáxia") e a respondem através de parâmetros extraídos do ajuste dos dados (i.e., da síntese espectral). O mapeamento do espaço de observáveis em um espaço de parâmetros envolve uma série de questões matemáticas e estatísticas complexas. Mesmo assim, este método é essencialmente calcado em princípios físicos.

Outros métodos são de natureza mais puramente matemática, buscando no espaço de dados observados estruturas e correlações que possivelmente revelem (ou pelo menos nos dêem pistas sobre) os fenômenos subjacentes. Aqui, a pergunta é bem definida em um sentido matemático, e o desafio é interpretar a resposta em termos físicos. Sánchez Almeida et al. (2010), por exemplo, categorizam espectros de galáxias do SDSS/DR7 em termos de clusters formados pelas distâncias euclidianas entre os pontos formados pelo espectro de cada galáxia em um espaço de dimensão N_λ . Cada cluster desse foi caracterizado por um espectro médio chamado de *K-mean* e é usado para posterior comparação com outros espectros. Um método bem mais popular e tradicional nessa mesma linha é a análise de componentes principais (PCA - *Principal Component Analisys*).

A técnica de PCA é simples, não-paramétrica, e nos ajuda a extrair informações de conjuntos de dados com muitas variáveis. Resumidamente, a PCA parte de uma tabela de N linhas (objetos) por M colunas (contendo propriedades observadas para cada objeto) e reorganiza as colunas transformando as propriedades originais em combinações delas mesmas de forma que as diferenças (a variância) entre os objetos esteja concentrada nas primeiras colunas, reduzindo assim dimensionalidade do problema (mais detalhes em 3.1). A tabela pode conter quaisquer tipos de dados, como cores, tamanho, dispersão de velocidade, luminosidade, etc. No caso de espectros de diferentes galáxias, as propriedades são as medidas de fluxo para cada comprimento de onda: $F_{gl} = F_g(\lambda_l)$, com $l = 1 \dots M$, para cada galáxia $g = 1 \dots N$.

Hoje, PCA é utilizada exaustivamente em várias áreas de conhecimento, principalmente

em reconhecimento de padrões, computação visual, filtragem e compactação de dados (Kamruzzaman et al. 2010; Borcea et al. 2012). Podemos ver exemplos também em medicina (Balakrishnan et al. 2013). Na astrofísica o PCA passeia por diversas ramificações da área. A luz de um objeto até nossos telescópios sofre influência de ruído e é afetada pela atenuação e avermelhamento por poeira, contaminação ótica através de objetos que estejam no mesmo campo, entre outros. Os próprios instrumentos também geram diversas assinaturas indesejadas. Por todos esses motivos fica claro que os dados geralmente necessitam de uma boa filtragem. Junto com outras técnicas (tomogramas, wavelets, Fourier), o PCA vem sendo muito utilizado para filtragem de dados, principalmente cubos de dados advindos de IFS que possuem muitas dessas assinaturas instrumentais (Riffel et al. 2011). Outro exemplo de uso de PCA aparece no artigo IV do grupo SEAGal/STARLIGHT (Mateus et al. 2007) auxiliando no estudo da dependência ambiental de algumas propriedades físicas (idade estelar média ponderada pela luz, massa estelar, metalicidade estelar e razão massa/luminosidade) em uma amostra de galáxias do *SDSS DR4*. Em Chen et al. (2012) é criada uma biblioteca com 25 mil modelos de espectros de galáxias com diferentes idades, metalicidades, velocidade de dispersão, SFH (*star formation history*), extinção por poeira e aplicado PCA em cima dessa biblioteca (ver Figura 1.1). Usando uma minimização quadrática encontram quais os coeficientes e qual o número ótimo de PCs que melhor estimam os parâmetros físicos dos espectros modelo. Então projetam os espectros observados pelo SDSS DR7 (Abazajian et al. 2009) e pelo Baryon Oscillation Spectroscopic Survey (BOSS Ahn et al. 2012), atribuindo um sentido físico a cada PC. Os trabalhos de **!○j○!** ??? ... Wild et al ... exemplificam a aplicacao de PCA a espectros do SDSS.

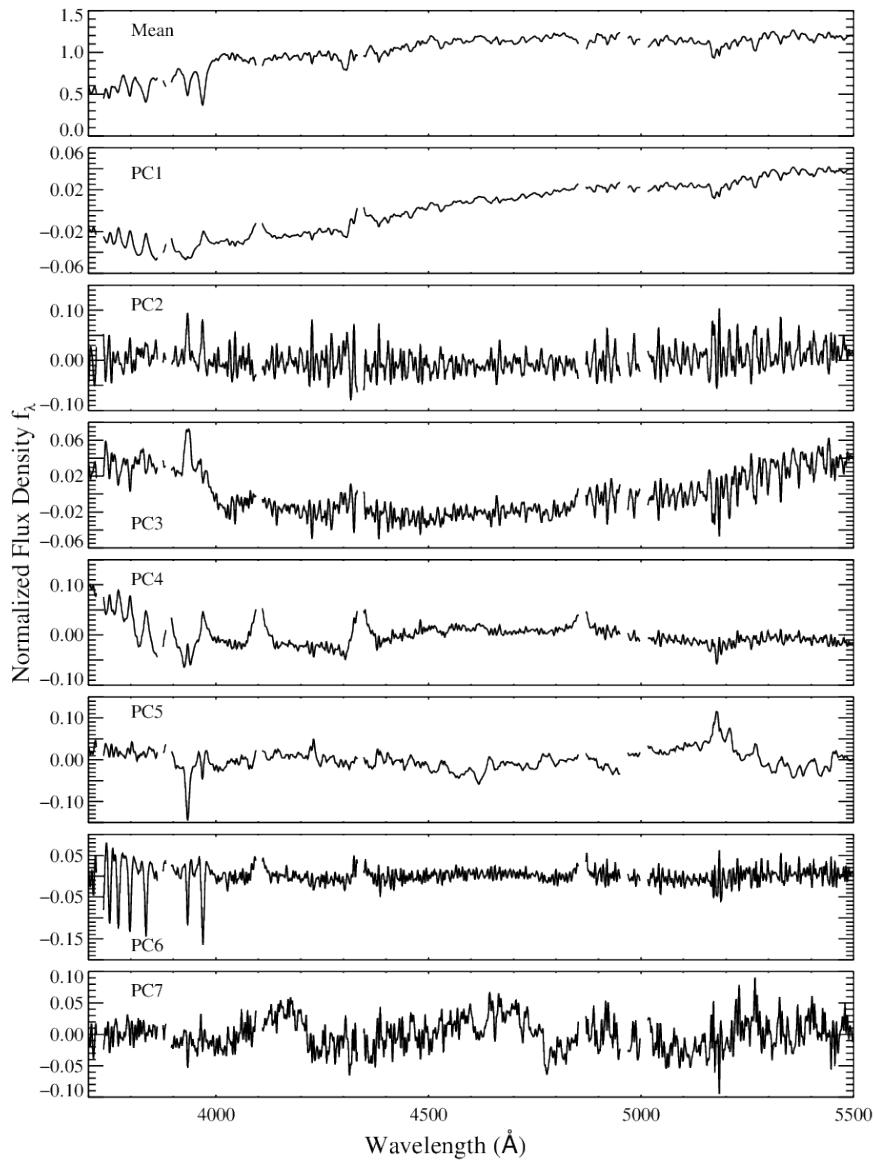


Figura 1.1: De cima para baixo: O espectro médio da biblioteca de espectros modelos seguido dos sete primeiros autoespectros da análise PCA. Retirado de Chen et al. (2012).

1.2 A nova ferramenta - Tomografia PCA

Todos exemplos acima relatam análises de espectros integrados de galáxias, ie., espectros que contém toda (ou pelo menos boa parte) da luz da galáxia inteira. Análises como a síntese espectral ou PCA comparam galáxias inteiras umas com as outras, ordenando-as em termos de propriedades físicas (e. g., a massa estelar ou a idade média das estrelas) ou matemáticas (e.g, suas componentes principais). Existem também exemplos de análises PCA ou de síntese espectral baseadas em espectros nucleares (e.g., Trager et al. 2000; Cid Fernandes et al. 2004). Assim como espectros integrados, tais estudos não possuem resolução espacial, ou seja, cada galáxia é representada por apenas um espectro.

Hoje, com o uso de painéis de fibras óticas apontadas para as galáxias, temos os *surveys* de IFS (*Integral Field Spectroscopy*) onde passamos a reunir de dezenas até milhares de espectros cobrindo um *FoV* de toda uma galáxia. Isso permite então que obtendo então uma “dimensão” no espaço também nos dados. Assim temos para cada píxel (duas dimensões espaciais) um espectro (uma dimensão espectral), formando assim um cubo de *spaxels*². Um pioneiro nessa produção em massa de dados é o *Calar Alto Legacy Integral Field spectroscopy Area survey*³ (CALIFA; Sánchez et al. 2012), produzindo cerca de quatro mil espectros por galáxia observada. Outros *mega-surveys* IFS estão por vir (veja a seção 2.1).

Com esses cubos de dados, cada galáxia passa a ser uma grande amostra de espectros, de modo que o que antes era feito galáxia a galáxia pode agora ser feito para diferentes regiões de uma mesma galáxia. Podemos assim executar o STARLIGHT para cada *spaxel* e então obtemos propriedades físicas das populações estelares em função da posição na galáxia. Este tipo de análise vem sendo realizado por pesquisadores do IAA (Enrique Pérez, Rosa González Delgado, Rubén García-Benito) e da UFSC (Roberto Cid Fernandes, André Luiz de Amorim, Natália Vale Asari). Aspectos técnicos e incertezas são discutidos em !ojô! (Cid Fernandes et al. 2013, 2014, CF13 e CF14 daqui pra frente), enquanto (Pérez et al. 2013), e (González Delgado et al. 2013) apresentam resultados astrofísicos dessa análise. Neste trabalho exploraremos os mesmos dados através da técnica de PCA.

A aplicação de PCA a cubos de dados foi realizada pela primeira vez por Steiner et al. (2009, S09 daqui pra frente), que cunharam o termo Tomografia PCA. Com essa nova técnica temos, além da análise das componentes principais, uma imagem formada pela matriz de covariância (as questões matemáticas serão abordadas no Capítulo 3). Assim podemos saber

²spectral pixels

³<http://www.caha.es/CALIFA/>

o peso, ou relevância, de cada componente principal (daqui pra frente PC) relacionada a uma posição na galáxia, combinando as virtudes de espectroscopia com as de imageamento.

Além de S09, outros exemplos de uso de Tomografia PCA podem ser encontrados em Riffel et al. (2011) e Ricci et al. (2011). Cumpre salientar que todas essas aplicações dessa nova técnica são baseadas em dados que cobrem apenas a **!○j○!** região central das galáxias, $3'' \times 5''$ ou Esta é naturalmente uma região rica em fenomenologia e portanto interessante de ser estudada com IFS, como ilustram os próprios resultados dos artigos supracitados **!○j○!** (ver tb sec ???).

1.3 Este trabalho

Através da colaboração do Grupo de Astrofísica da Universidade Federal de Santa Catarina (GAS-UFSC) com o grupo de pesquisadores do projeto CALIFA temos a oportunidade de trabalhar com os dados de IFS das galáxias observadas por esse projeto, que ainda está em andamento. O seu primeiro *Data Release*⁴ (Husemann et al. 2013, DR1) possui 100 objetos e por volta de 400 mil espectros. A previsão é que ao término do projeto serão observadas até 600 objetos. Embora outros surveys de IFS estão completos ou em andamento (ver Seção 2.1), o CALIFA é o que podemos chamar de *estado da arte* em surveys de espectroscopia de campo integrado (IFS).

Como já citado, pesquisadores do IAA e da UFSC vêm aplicando o STARLIGHT aos cubos do CALIFA e com isso mapeando as propriedades das populações estelares nessas galáxias. Mais de 300 cubos de dados já foram analisados, embora os resultados publicados até agora se restrinjam a pouco mais que 100 galáxias, de espirais tardias até elípticas, incluindo alguns sistemas em interação. Ferramentas desenvolvidas nesses trabalhos anteriores serão utilizadas nessa dissertação, como veremos mais adiante.

Neste trabalho aplicaremos a técnica de Tomografia PCA a alguns cubos de dados do CALIFA. O objetivo é pura e simplesmente ver o que se obtém! Como bem colocado por Steiner et al. (2009), o grande problema da PCA é que você tem a resposta, mas não sabe exatamente qual é a pergunta! Este é, portanto, um estudo eminentemente exploratório, motivado pelo sucesso dos trabalhos de Steiner e colaboradores em suas aplicações dessa nova ferramenta de análise de cubos dados.

É importante notar, no entanto, que existe uma grande diferença entre esses trabalhos

⁴http://www.caha.es/CALIFA/public_html/?q=content/califa-dr1

anteriores e esse estudo. Nossos cubos de dados cobrem uma escala completamente diferente. De fato, o CALIFA foi desenhado para observar galáxias inteiras. O campo do instrumento utilizado é de aproximadamente 1 arcmin². Em comparação com os estudos de Steiner e colaboradores, temos uma resolução espacial muito "pior", mas em compensação mapeamos estruturas muito diferentes, como o bojo, disco, braços e regiões HII a muitos kpc do núcleo. Como veremos, esta diferença nos levou a tratar os dados de formas que diferem daquelas empregadas por Steiner e seus colaboradores. De fato, até onde temos conhecimento, este trabalho é o primeiro a aplicar a técnica de tomografia PCA à galáxias interias.

1.3.1 Organização deste trabalho

No seguinte capítulo apresentamos de forma resumida o *survey* CALIFA e dois pipelines importantes para a organização dos dados utilizados neste trabalho: PyCASSO e QBICK.

O terceiro capítulo descreve matematicamente a técnica de PCA e a Tomografia PCA, bem como sua utilização no presente trabalho.

No quarto capítulo discutimos diferentes *configurações*⁵ de execução da PCA, juntamente com suas implicações aos resultados das análises. As diversas escolhas relativas às etapas de pré-processamento dos dados são discutidas de forma exemplificada.

Com todo o arcabouço teórico em mãos, no capítulo cinco temos um estudo de caso para 8 objetos do CALIFA, escolhidos para abranger "de tudo um pouco". São galáxias 4 espirais, 2 *early-type* e 2 objetos compostos (mergers). Como forma de comparação com os resultados da síntese de populações estelares executadas pelo STARLIGHT nesses objetos também é feito uma espécie de engenharia reversa através de correlações e comparações.

Por fim, temos as conclusões e perspectivas futuras deste trabalho e da nossa colaboração com o projeto CALIFA no sexto e último capítulo.

⁵Configurações aqui é usado apenas como uma maneira de tipificar diferentes pré-processamentos ou conjuntos dos dados antes da execução da PCA.

Capítulo 2

O projeto CALIFA e o *pipeline* PyCASSO

Nesse cenário de crescimento exponencial na quantidade de dados alavancados por novas tecnologias surge o CALIFA: um projeto que está modificando nossa maneira de ver e pensar as galáxias no nosso universo de forma que entendamos melhor a nossa também. Na primeira seção deste capítulo iremos apresentar o survey CALIFA e suas diferenças sobre os surveys atuais.

Com a enorme quantidade de dados obtidos, resultado direto de um projeto de ciência de ponta, vem também a dificuldade da organização dos dados para futuras investigações científicas. Dois *pipelines* nos ajudam na preparação dos dados que iremos usar neste trabalho, qbick e PyCASSO. O qbick prepara os dados para a síntese de populações estelares e o PyCASSO organiza a saída da síntese, que é executada para cada espectro dentro da galáxia. Na segunda seção apresentaremos ambos, mostrando as peculiaridades e técnicas empregadas sobre os dados através do qbick e da descrição e alguns exemplos de utilização do PyCASSO.

2.1 O survey CALIFA

No sul da Espanha, mais precisamente em *Sierra de Los Filabres* (Andalucía), está situado Observatório Astronômico Hispano-Alemão de *Calar Alto*. O projeto CALIFA está sendo possível através de observações pelo maior de seus 3 telescópios (3.5m) ao longo de 250 noites. Em comparação com o *SDSS*, o CALIFA terá a mesma ordem de número de espectros para estudo ($\sim 10^6$), mas, apesar de um número muito menor de galáxias, graças ao IFU será o

com melhor completeza por objeto. Existem alguns poucos surveys IFU e todos com, além de poucos objetos e FoV menor, focos de estudo mais estreitos, dificultando o legado do survey para outras pequisas científicas mais abrangentes (SAURON; de Zeeuw et al. 2002, região central de 72 galáxias com $z < 0.01$) (PINGS; Rosales-Ortega et al. 2010, algumas galáxias muito próximas (~ 10 Mpc) e o estudo atual de 70 (U)LIRGs com $z < 0.26$) (VENGA; Blanc et al. 2010, 30 galáxias espirais). O CALIFA foi concebido para que seu legado seja abrangente, possibilitando diversos tipos de estudos em diversas áreas. Outros surveys IFU ainda estão por vir, como SAMI (Croom et al. 2012) e MaNGA¹

A amostra-mãe do projeto comporta 939 galáxias (das quais ~ 600 serão observadas) com *redshifts* entre $0.005 < z < 0.03$ que distribuídas cobrem o diagrama cor-magnitude com $M_r < -18$ (Figura 2.1) em uma ampla variedade de tipos morfológicos, massa em estrelas, condições do gás ionizado. Para melhor aproveitar o *FoV* (*field-of-view*²) do instrumento de IFU é feito também um corte em dimensão ($\sim 1'$ em diâmetro). No telescópio usado para o projeto está instalado o equipamento Potsdam Multi Aperture Spectrograph (PMAS; Roth et al. 2005) no modo PPAK (Verheijen et al. 2004; Kelz et al. 2006) formando um espectrofotômetro de campo integrado com um *bundle* de 382 fibras (Figura 2.2), das quais, 331 são para observação dos objetos, outras 36 que fazem amostras para a sutração do céu e outras 15 para calibração. As *science fibers* (331) cobrem um campo de visão hexagonal de $74'' \times 64''$ que, através de uma técnica de três pontos de *dithering* torna possível a observação de 100% do campo.

Os dados são reduzidos utilizando a CALIFA Pipeline versão 1.3c, descrito em Husemann et al. (2013, ap. A). Com as 331 *science fibers* e o processo de *dithering* em 3 posições, temos 993 espectros independentes por objeto observado. Com a reamostragem do cubo de espectros de forma que cada spaxel tenha área de $1'' \times 1''$ obtém-se mais de 4000 espectros por galáxia. Isto implica que cada píxel do cubo final não seja completamente independente, ou seja, numa automática correlação entre os píxeis vizinhos. Os espectros vêm em duas configurações de rede: a V500 cobrindo de ~ 3700 até 7000 Å com resolução de ~ 6 Å de largura à meia altura (FWHM) e a V1200 (~ 3650 – 4600 Å FWHM ~ 2.3 Å). A cobertura espectral da V500 é ideal para os propósitos de ciência feita pelo STARLIGHT, mas por problemas com *vignetting* com a parte azul dessa configuração, os dados são reamostrados numa combinação das duas redes, criando cubos que chamamos de COMBO: A parte com $\lambda < 4600$ Å vem da V1200 e o resto da V500. Nessa montagem os espectros da V1200 foram suavizados à mesma resolução espectral da V500 (FWHM = 6 Å). Como outros cubos do CALIFA, os cubos COMBO provêm, para cada posição e comprimento de onda, o fluxo ($F_{\lambda,x,y}$), seu erro ($\epsilon_{\lambda,x,y}$) e uma flag ($b_{\lambda,x,y}$) que

¹<http://www.sdss3.org/future/manga.php>

²campo de visão

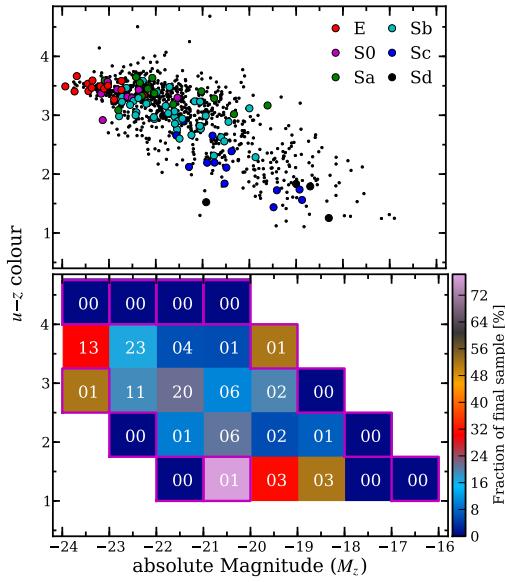


Figura 2.1: Distribuição das galáxias do CALIFA no diagrama cor magnitude $u-z$ vs. M_z . *Painel superior*: Em pontos pretos estão as galáxias pertencente a amostra-mãe e em cores as galáxias presente no CALIFA DR1. As diferentes cores representam os diferentes tipos morfológicos. *Painel inferior*: A fração de galáxias observadas pelo DR1 em relação a amostra-mãe. Retirado de Husemann et al. (2013).

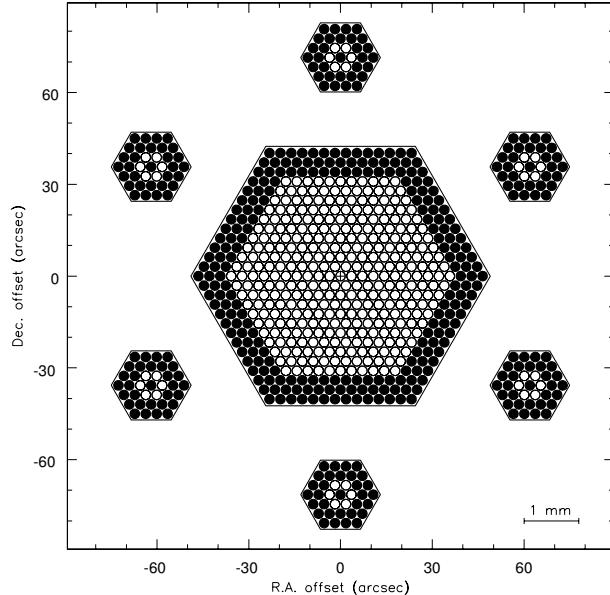


Figura 2.2: Este é o esquema com o *bundle* hexagonal com as 331 fibras de observação e mais 36 de amostra de céu. Retirado de Verheijen et al. (2004).

sinaliza pixeis defeituosos.

Nesse trabalho usaremos exclusivamente dados dos cubos COMBO. Por alguns motivos que veremos na próxima seção não analisarermos exatamente esses dados em sua forma original, mas os provenientes de duas outras pipelines desenvolvidas por nossa equipe.

2.2 As pipelines QBICK e PyCASSO

Em CF13 temos a descrição de duas outras pipelines usadas na análise dos cubos COMBO: O QBICK e o PyCASSO. Abaixo revisamos essas pipelines, ambas muito importantes nesse trabalho.

2.2.1 QBICK

A primeira é a QBICK, desenvolvida por Rubén García-Benito, do IAA, e explicada em detalhe em CF13. Resumidamente, esta pipeline:

1. Cria máscaras espaciais que removem regiões externas de baixo S/N, e objetos indesejáveis no *FoV* (estrelas ou galáxias).
2. Introduz flags espectrais que indicam janelas contendo linhas telúricas, cuja remoção imperfeita pode afetar os espectros das regiões de menor fluxo.
3. Desloca o espectro para o referencial de repouso (usando o *redshift* calculado dentro dos 5'' centrais da galáxia) e o reamostra a 2 Å.
4. Define zonas de Voronoi, tal que o espectro total dessas zonas tem uma razao S/N ≥ 20 em uma janela espectral de 90 Å de largeura e centrada em 5635 Å, e extrai os espectros correspondentes.

A Figura 2.3 ilustra as máscaras espectrais utilizadas, tomando como exemplo a zona 399 para a galáxia NGC 2916 (CALIFA 277). O painel superior mostra o espectro original. No painel do meio as janelas mascaradas pelo QBICK são removidas do espectro. Essas janelas correspondem a regioes contendo linhas telúricas³, as quais dificilmente sao perfeitamente removidas na reducao. Pixeis identificados como problemáticos durante a reducao dos dados

³Linhas de emissão ou absorção referentes à atmosfera.

sao igualmente mascarados. Em nossa análise fazemos uma estatística com todos os *bad pixels* de cada cubo e removemos qualquer um que esteja presente em mais de 5% dos espectros. Cabe aqui lembrar que todos os espectros precisam ter os mesmos pontos em comprimento de onda pois precisamos construir a matriz de covariância.

Além dos *bad pixels* e linhas telúricas, podemos mascarar regiões desnecessárias para determinada investigação científica. Nossa foco é o estudo das populações estelares, e portanto optamos por mascarar as linhas de emissão. Escolhemos mascarar as seguintes janelas: H ϵ : de 3960 a 3980 Å; H δ : de 4092 a 4112 Å; H γ : de 4330 a 4350 Å; H β : de 4848 a 4874 Å; [O III]: de 4940 a 5028 Å; He I e NaD: de 5866 a 5916 Å; H α e [N II]: de 6528 a 6608 Å; S II: de 6696 a 6752 Å. Essas máscaras são na verdade introduzidas em uma etapa posterior ao QBICK, quando os espectros sao processados com o STARLIGHT. Também nessa etapa os espectros são cortados entre 3800 e 6850 Å. O espetro aps todas máscaras pode ser visto no painel inferior da Figura 2.3.

O efeito do agrupamento em zonas de Voronoi é ilustrado na Figura 2.4 para a galáxia NGC 2916 cujos ??? pixeis úteis (ie, não mascarados espacialmente) são agrupados em $N_z = 16??$ zonas. Como esperado, o efeito é maior nas partes externas. Vale aqui frisar que a grande maioria das zonas comporta um pixel apenas. Para a galáxia NGC 2916, por exemplo, 93% das zonas possuem apenas 1 pixel e ~ 3% com mais de 10 pixeis agrupados, mesmo com a imposição de relaçao S/N ≥ 20 em $\lambda = 5635$ Å, nosso comprimento de onda de referêcia (veja Tabela 2.1, que também mostra dados para outras galáxias estudadas esse trabalho). Portanto o cubo original é transformado numa matriz de zonas e comprimentos de onda.

A motivação para a "zonificacao"do cubo é garantir espetros de qualidade boa o suficiente para análise com o STARLIGHT. Nossa análise PCA poderia ser realizada sobre os cubos de dados originais, mas a realizaremos sobre as zonas produzidas pelo QBICK. A razao para isso é que queremos comparar os resultados de nossa análise com os parâmetros extraídos pelo STARLIGHT, o que servirá de auxílio a interpretacao dos resultados da PCA.

Todas informações e pré-processamentos provenientes do QBICK são herdadas pelo PyCASSO e já estão contidas em seus cubos de espetros.

2.2.2 PyCASSO

Os espetros extraídos pelo QBICK são processados pelo STARLIGHT, produzindo, no caso de CALIFA 277, 1639 arquivos ASCII como output. Dado o grande número de informações

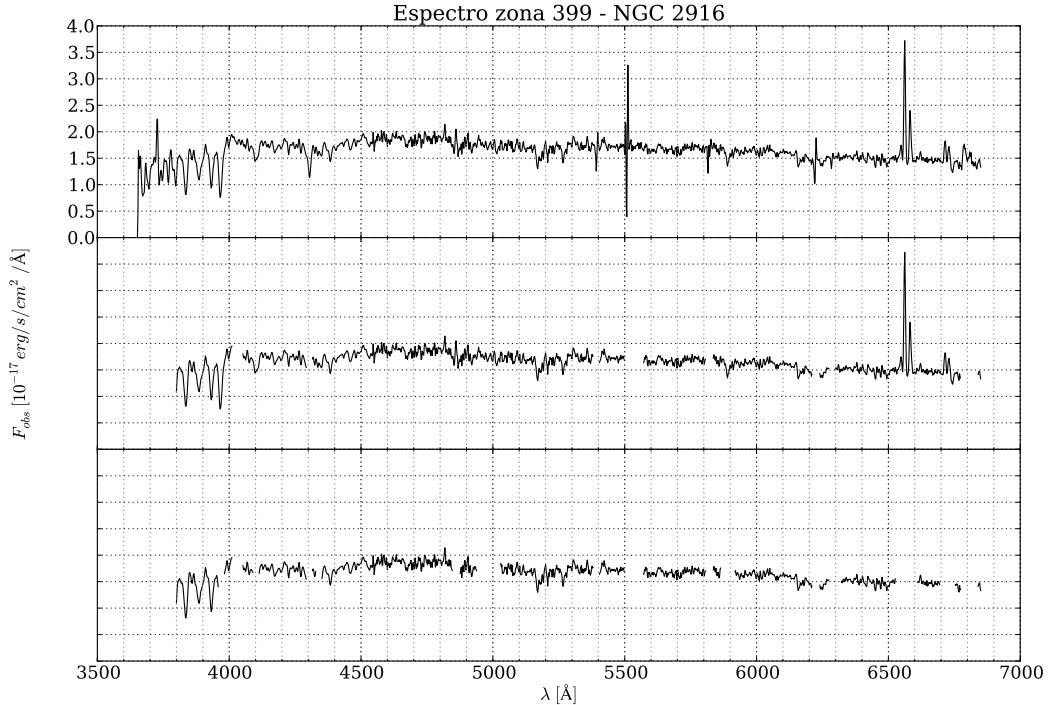


Figura 2.3: Espectro da zona 399 da galáxia NGC 2916 (CALIFA 277). Acima está o espectro completo. No segundo vemos o espectro com linhas telúricas e bad-pixels removidos, além do limite de intervalo em comprimento de onda de 3800 a 6850 Å. No espectro mais abaixo, além das partes removidas no segundo, estão removidas também as mesmas linhas de emissão mascaradas na síntese de populações estelares.

Figura 2.4: !@j@! FIG 1 DE CF13..... (a) imagend sdss... (b) blabla ... (c) ... (a) SDSS stamp for NGC 2916 (CALIFA 277). As in other images throughout this paper, the circles mark $R = 1$ and 2 Half Light Radius (HLR). (b) CALIFA image in the rest-frame 5635 ± 45 Å interval after applying the spatial mask. (c) Map of the S/N at 5635 Å. (d) S/N map after the Voronoi binning. (e) Number of spaxels in each Voronoi zone. Note that no spatial binning ($N_z = 1$) was necessary throughout the inner ~ 1 HLR. (f) Percentage of bad pixels in the 3800–6850 Å range. North is up and east is towards the left in all images. Retirado de ?).

Tabela 2.1: Relação entre números de pixels por zona nas galáxias do CALIFA utilizadas neste trabalho. N_z representa o número de zonas. N_1 é o numero de zonas com 1 pixel apenas e N_{10} aquelas que possuem mais de 10 pixels por zona.

Nome da galáxia	CALIFA ID	Hubble Type	N_z	N_1	N_1 / N_z	N_{10}	N_{10} / N_z
NGC 0001	K0008	Sbc	1132	1077	0.95	40	0.04
NGC 0776	K0073	Sb	1733	1628	0.94	61	0.04
NGC 1167	K0119	S0	1879	1771	0.94	50	0.03
NGC 2623	K0213	Scd	561	530	0.94	19	0.03
NGC 2916	K0277	Sbc	1638	1528	0.93	53	0.03
NGC 4210	K0518	Sb	1938	1847	0.95	38	0.02
ARP 220	K0802	Sd	1157	1103	0.95	39	0.03
NGC 6515	K0864	E3	887	811	0.91	44	0.05

contidas em cada um desses arquivos (história de formacao estelar em luz e massa, extincao, parametros conematicos, figuras de merito do ajuste, etc.), necessitamos de um *pipeline* que faça a organização de todos os dados para que cálculos e gráficos de mais variadas dificuldade sejam fácil para que até um programador de nível iniciante possa fazer. André L. de Amorim, juntamente com outros colaboradores de nosso grupo e do projeto CALIFA construiu o PyCASSO (*Python CALIFA STARLIGHT Synthesis Organizer*) (? , cap. 4) que faz a organização dos dados que vêm do QBICK e das etapas iniciais de reducao juntamente com a síntese de populações estelares feitas com o STARLIGHT, facilitando, e muito, o trabalho de quem usa estes dados. Sem a ajuda deste organizador, este trabalho seria muito mais difícil, além de infactível no tempo de um mestrado.

O PyCASSO é uma biblioteca desenvolvida em Python para organizar os dados da síntese feita pelo STARLIGHT. A versão usada neste trabalho é a 0.9.3⁴. Para um acesso mais rápido e reutilizável do código e dos dados em qualquer ambiente, organiza os cubos em formatos FITS ou HDF5. Em outra camada, várias matrizes e cubos são armazenados para acesso com nomes próprio (um exemplo, `popx`, representa a fração de luz distribuída pelas populações estelares da base usada na síntese espectral com o STARLIGHT) de forma que a programação exploratória não precise se preocupar com as características de cada formato de armazenamento de dados. Por fim, existe uma camada construída para análise, com funções que retornam a indexação de cada zona para um par (x, y) , cálculos de perfis radiais e azimutais, geometria, entre outras rotinas. Um programador facilmente pode adicionar mais rotinas como essas.

⁴<http://minerva.ufsc.br/~andre/PyCASSO-0.9.3/>

```

1 # Carregar arquivo FITS com os dados.
2 from pycasso import fitsQ3DataCube
3 K = fitsQ3DataCube('K0277_synthesis_suffix.fits')
4
5 # Acessar a idade media ponderada pela luminosidade.
6 at = K.at_flux_z
7
8 # Calcular a idade media da galaxia.
9 at_total = (at * K.Lobn_z).sum() / K.Lobn_z.sum()
10 print 'Idade media da galaxia K0277: %.2f' % at_total

```

Figura 2.5: Exemplo de acesso aos cubo de dados por arquivo FITS e o cálculo da idade estelar média de uma galáxia.

Todo os programas desenvolvidos nessa dissertação foram escritos de forma compatível com o PyCASSO.

!○j○! NAO SEI ONDE ENCAIXAR ISSO... A PRIMEIRA FRASE ATEH QUE PODERI SER UTILIZADA LAH NA INTRO DO CAP. ... A beleza da espectroscopia de campo integral é poder unir imageamento e espectroscopia, obtendo-se assim a galáxia vista como um cubo de dados (x, y, λ); para cada λ temos uma imagem, e para cada par de ascenção reta e declinação (x, y) temos um espectro. !○j○! Pedir uma imagem chique de cubo para o André. Apesar de cada espectro poder ser analizado individualmente, nos cubos do CALIFA existe uma correlação entre pixels vizinhos devido a !○j○! seeing do céu, ao tamanho das fibras (??? de diâmetro) e ao processo de observação e reducao. Huseman et al 2013 estimam que a resolução espacial do CALIFA é de 3.7 arcsec (correspondendo tipicamente ??? kpc para as galáxias da amostra), mas estudos posteriores indicam que ela é um pouco melhor que isso.

2.2.3 Exemplos de utilização do PyCASSO

Ler um arquivo FITS é fácil com o PyCASSO, assim como o acesso aos dados. Na Figura 2.5 temos um exemplo de leitura de arquivo FITS e um cálculo da idade estelar média da galáxia a partir da idade média ponderada pela luminosidade, por zona (at_flux_z). A idade estelar média é calculada usando a expressão $\langle \log t \rangle_L^{gal} = \sum_z \langle \log t \rangle_{L,z} L_z / \sum_z L_z$ onde L_z é a luminosidade e $\langle \log t \rangle_L$ é a idade estelar média, ambas por zona. A passagem de coordenadas $z \rightarrow (x, y)$ exemplificada na Figura 2.7 é feita através da função de zoneToYX dentro do PyCASSO.

```

1 # Carregar arquivo FITS com os dados.
2 from pycasso import fitsQ3DataCube
3 K = fitsQ3DataCube('K0277_synthesis_suffix.fits')
4
5 # Converter zonas para imagem.
6 at_image = K.zoneToYX(K.at_flux__z, extensive=False)
7
8 # Desenhar o mapa.
9 import matplotlib.pyplot as plt
10 plt.imshow(at_image, origin='lower', interpolation='nearest')
11 plt.xlabel('Pixels')
12 cb = plt.colorbar()
13 cb.set_label(r'\langle \log t \rangle_L [anos]')
14 plt.title(r'\langle \log t \rangle_L z')

```

Figura 2.6: Programa para desenhar o mapa de idade estelar média ponderada pela luminosidade.

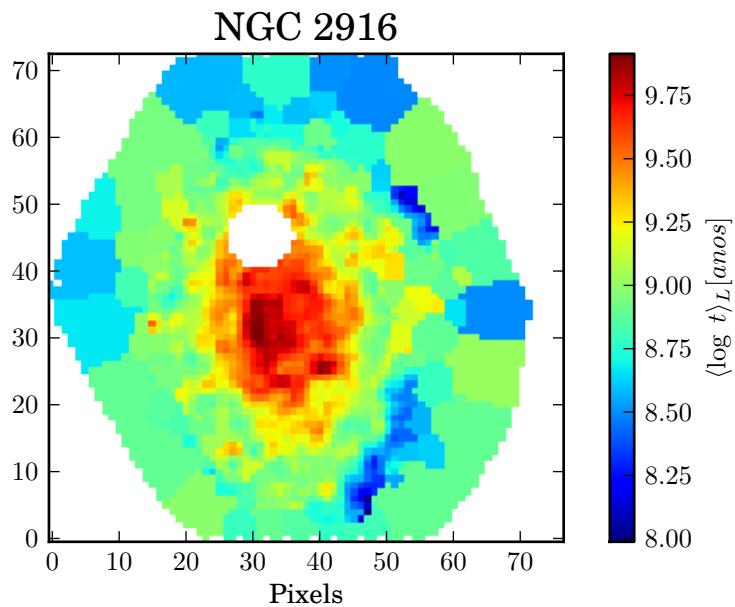


Figura 2.7: Mapa de idade estelar média ponderada pela luminosidade da galáxia NGC 2916 (CALIFA 277) gerado pelo programa da Figura 2.6. !ôjo! Será que não vale a pena botar umas imagens de AV e ev0, p ex, para ilustrar mais? Pode servir de contraponto aos tomogramas. Acho uma boa ideia! Enfim, pensa ai nessa possibilidade.

```

1 # Carregar arquivo FITS com os dados.
2 from pycasso import fitsQ3DataCube
3 K = fitsQ3DataCube('K0277_synthesis_suffix.fits')
4
5 # Converter zonas para imagem.
6 at_image = K.zoneToYX(K.at_flux__z, extensive=False)
7
8 # Calcular o perfil radial.
9 bins = np.arange(0, 26, 1)
10 bin_center = (bins[1:] + bins[:-1]) / 2.0
11 at_rad = K.radialProfile(at_image, bins, rad_scale = 1.0)
12
13 # Desenhar o perfil.
14 import matplotlib.pyplot as plt
15 plt.xlabel('radius [arcsec]')
16 plt.ylabel(r'\langle \log t \rangle_L [anos]')
17 cb = plt.colorbar()
18 plt.plot(bin_center, at_rad)

```

Figura 2.8: Programa para desenhar o perfil radial da idade estelar média ponderada pela luminosidade.

Perfis radiais ou axiais também são importantes para o estudo de diversas propriedades galácticas. Podemos ver na Figura 2.8 (e em sua imagem gerada 2.9) um exemplo de perfil radial executado através da função `radialProfile` do PyCASSO.

Já existem vários artigos utilizando o PyCASSO (Cid Fernandes et al. 2013, 2014; Pérez et al. 2013; González Delgado et al. 2013), também há alguns usando indiretamente (Husemann et al. 2013; Iglesias-Páramo et al. 2013) e três teses de doutorado em curso (uma na USC e duas no IAA). São mais de 10 pessoas utilizando a biblioteca em nossa colaboração.

Usando o `matplotlib`⁵ podemos fazer gráficos facilmente acessando as matrizes e vetores do PyCASSO, como pode ser visto no programa na Figura 2.6 e sua imagem gerada (Figura 2.7). Os cálculos necessários para o PCA e também para a Tomografia PCA se tornam simples contas usando pacotes matemáticos de python. Todo software desenvolvido nesta dissertação (por ora apelidado de "PCAlifa") foi feito de forma compatível com o PyCASSO, visando futuramente integrá-lo como um novo módulo dessa ferramenta.

⁵<http://matplotlib.org>

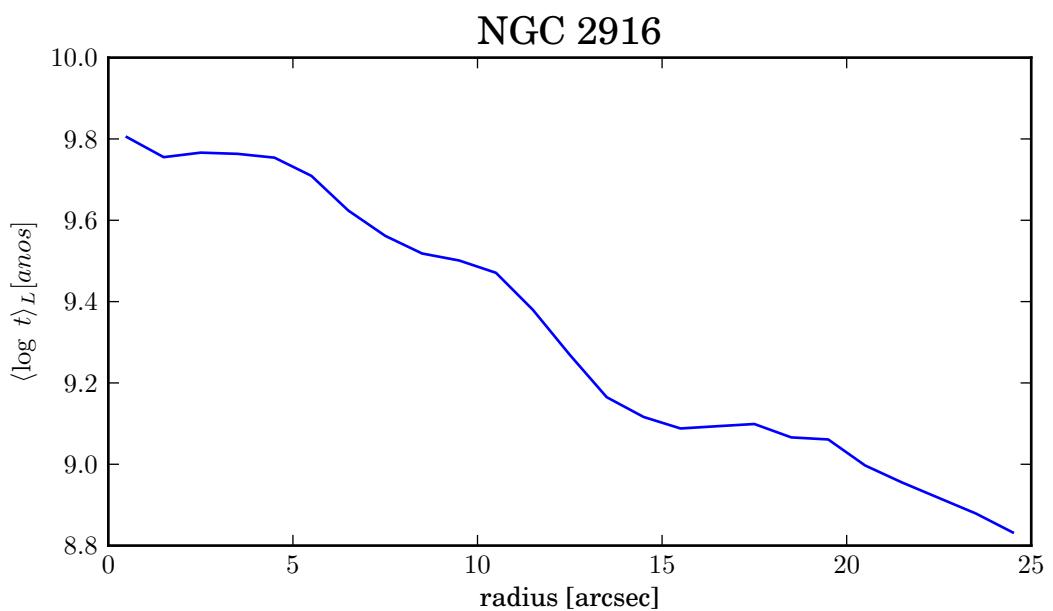


Figura 2.9: Perfil radial da idade estelar média ponderada pela luminosidade da galáxia NGC 2916 (CALIFA 277) gerado pelo programa da Figura 2.8.

Capítulo 3

PCA e Tomografia PCA

De medidas fisiológicas, como pulsação e respiração, até reconhecimento de padrões em sistemas complexos como reconhecimento facial e criptografia, passando por compactação de imagem, neurociência e redução de ruídos em dados, podemos ver atuação de técnicas de PCA. Neste capítulo revisamos os fundamentos matemáticos do PCA e de sua versão para cubos de dados (Tomografia PCA). **Essa intuito sim está perfeita :-)**

3.1 Principal Component Analysis

Baseada em encontrar os eixos com maiores variâncias em um conjunto de variáveis (no nosso caso, fluxos por lambda e por zona espacial), a técnica de PCA vem sendo de grande utilidade quando o assunto é estatística com muitas variáveis. Através de operações relativamente simples usando álgebra linear, é feita uma rotação na base de dados original, gerando uma nova base ortonormal não correlacionada através de um conjunto de autovalores e autovetores.

Existem diversas formas de se calcular essa base final. A prova matemática que você pode obter essa base é feita através de multiplicadores de Lagrange, calculando os autovetores (\mathbf{e}_k) e autovalores (λ_k) **[!]** **! Melhor usar Λ_k pra não confundir com comprimento de onda...]** que maximizam o valor de $\mathbf{e}_k^T \cdot \mathbf{C}_{cov} \cdot \mathbf{e}_k$ (ver Eq. 3.4) sujeito à restrição de que um autovetor deve ser ortogonal a qualquer outro da base ($\mathbf{e}_i^T \mathbf{e}_j = 0$) e que todos devem ser normalizados ($\mathbf{e}_i^T \mathbf{e}_i = 1$) (Jolliffe 2002, p. 5-6). No caso de PCA com espectros, cada um pode ser representado como um ponto num espaço **[!]** N_λ -dimensional. Vários espectros formam uma nuvem de pontos nesse espaço. Assim, encontramos quais são os eixos mais significativos desse espaço em relação à variância, sujeitos às restrições acima. Fazemos esse cálculo

encontrando os autovetores e autovalores da matriz de correlação **[!jō! Covariancia?]** desse espaço (a matriz tem dimensão λ) usando a biblioteca científica SciPy¹ (3.1), que encontra os autovetores e autovalores da matriz de correlação **[!jō! Covariancia?]**.

3.1.1 PCA das galáxias do CALIFA

Conforme a Seção ?? vimos que o cubo de espectros das galáxias do CALIFA estão acessíveis no PyCASSO separados por zonas, muitas correspondendo a píxeis individuais e outras a conjuntos de pixeis (zonas de Voronoi). Os espectros observados estão armazenados em forma de uma matriz $(n \times m)$ **[!jō! Conferir notacao ... n = N_z?? m = N_\lambda??]** com n zonas e m comprimentos de onda (`f_obs`² no PyCASSO).

$$\mathbf{F}_{z\lambda} = \begin{bmatrix} f_{z_0\lambda_0} & f_{z_0\lambda_1} & f_{z_0\lambda_2} & \dots & f_{z_0\lambda_m} \\ f_{z_1\lambda_0} & f_{z_1\lambda_1} & f_{z_1\lambda_2} & \dots & f_{z_1\lambda_m} \\ f_{z_2\lambda_0} & f_{z_2\lambda_1} & f_{z_2\lambda_2} & \dots & f_{z_2\lambda_m} \\ \dots & \dots & \dots & \dots & \dots \\ f_{z_n\lambda_0} & f_{z_n\lambda_1} & f_{z_n\lambda_2} & \dots & f_{z_n\lambda_m} \end{bmatrix} \quad (3.1)$$

Calculamos então o espetro médio de uma galáxia

$$\langle \mathbf{F}_\lambda \rangle = \frac{1}{n} \sum_{i=0}^n f_{z_i\lambda} \quad (3.2)$$

e então subtraímos a média de todos os espetros

$$\mathbf{I}_{z\lambda} = \mathbf{F}_{z\lambda} - \langle \mathbf{F}_\lambda \rangle \quad (3.3)$$

para o cálculo da matriz de covariâncias usando um conjunto de dados com média zero:³

¹<http://scipy.org/>

²No PyCASSO está amostrado de forma transposta ($\lambda \times z$) a usada.

$$\mathbf{C}_{cov} = \frac{[\mathbf{I}_{z\lambda}]^T \cdot \mathbf{I}_{z\lambda}}{n - 1} \quad (3.4)$$

Vemos que a matriz de covariância possui dimensão $\lambda \times \lambda$. [!@j@! $m \times m?? N_\lambda \times N_\lambda ??$]

Agora calculamos os autovalores e autovetores da matriz de covariância. Neste trabalho usamos o nome autoespectro para designar esses autovetores pois são de uma matriz de covariâncias entre espectros de cada zona. Então ordenamos-os decrescentemente pelo valor de seus autovalores. Os autoespectros são as componentes principais (PCs) e os autovalores as respectivas variâncias. Isso feito, temos então o que necessitamos para iniciar o cálculo do Tomograma PCA.

Muitas figuras de PCs e suas utilizações e diferenças nos pré-processamentos serão mostradas nos Capítulos ?? e 5, juntamente com a Tomografia PCA e as comparações com os parâmetros físicos da síntese de populações estelares com o STARLIGHT.

```
1 # Carregar arquivo FITS com os dados.
2 from pycasso import fitsQ3DataCube
3 K = fitsQ3DataCube('K0277_synthesis_suffix.fits')
4
5 # Calcular o espectro medio de uma galaxia.
6 # K.f_obs tem dimensao (lambda, zona), portanto,
7 # fazemos o espectro medio de todas as zonas.
8 f_obs_mean_1 = K.f_obs.mean(axis = 1)
9
10 # Subtraímos a media
11 I_obs_zl = K.f_obs.transpose() - f_obs_mean_1
12
13 # Calcular a matrix de covariância
14 import scipy as sp
15 n = K.N_zone
16 dot_product = sp.dot(I_obs_zl.transpose(), I_obs_zl)
17 covMat_11 = dot_product / (n - 1.0)
18
19 # Calcular os autovalores e autovetores
20 w, e = linalg.eigh(covMat_11)
21
22 # Ordenar os autovetores decrescentemente pelo seu autovalor
23 S = sp.argsort(w)[::-1]
24 eigval = W[S]
25 eigvect = e[:, S]
```

Figura 3.1: Cálculo do procedimento completo de PCA para os espectros observados de uma galáxia do CALIFA usando o PyCASSO e a biblioteca científica de Python chamada SciPy. No final do código temos em eigval e eigvect os autovalores e autovetores ordenados em forma decrescente. **!OJO!** O TEXTO NAO FAZ REF A ESTA "FIG"!!

3.2 Tomografia PCA

Na técnica de PCA procuramos os autoespectros (PCs) da matriz de correlação, ordenados pela variância, formam uma base ($\mathbf{E}_{\lambda k}$) onde podemos projetar os nossos dados através da transformação:

$$\mathbf{T}_{zk} = \mathbf{I}_{z\lambda} \cdot \mathbf{E}_{\lambda k} \quad (3.5)$$

Projetamos nossa matriz de observáveis com a média subtraída ($\mathbf{I}_{z\lambda}$) na base das PCs ($\mathbf{E}_{\lambda k}$). Na posse dessa nova matriz transformada e de um mapa que leve de zona para uma par de coordenada ($z \rightarrow (x, y)$), podemos montar assim uma imagem. Cada imagem funciona como uma “fatia” de um cubo de dados expandido na nova base, assim formando a Tomografia PCA³, criada e assim batizada por Steiner et al. (2009), que em seu artigo faz um paralelo com fatias de um espaço tridimensional (tomograma do corpo humano, por exemplo) ou no espaço de velocidades (Tomografia Doppler). Cada “fatia” possui um autoespectro relacionado que, em conjunto, trazem novas perspectivas e ideias para a interpretação de ambos.

3.3 Exemplos de resultado da Tomografia PCA

Para ilustrar o potencial da tecnica de tomog PCA, ilustramos os resultados obtidos em 2 artigos.

3.3.1 Descoberta de linhas largas em um LINER

No artigo citado anteriormente, através do estudo dos autoespectros e suas respectivas imagens do núcleo da galáxia LINER (*Low Ionizations Nuclear Emission-line Region*) NGC 4736, foram encontradas evidências de linhas largas. Quando temos uma fonte que é capaz de produzir linhas largas no espectro é sinal da existência de um SMBH (*Super Massive Black Hole*). Cid Fernandes et al. (2004) mostram que, pelo menos em alguns casos, a subtração detalhada das populações estelares nos espectros ajuda a encontrar linhas largas mais fracas em Seyferts 2, que são aquelas que (por definição) possuem apenas linhas estreitas, ajudando assim na clas-

³<http://www.astro.iag.usp.br/~pcatomography/>

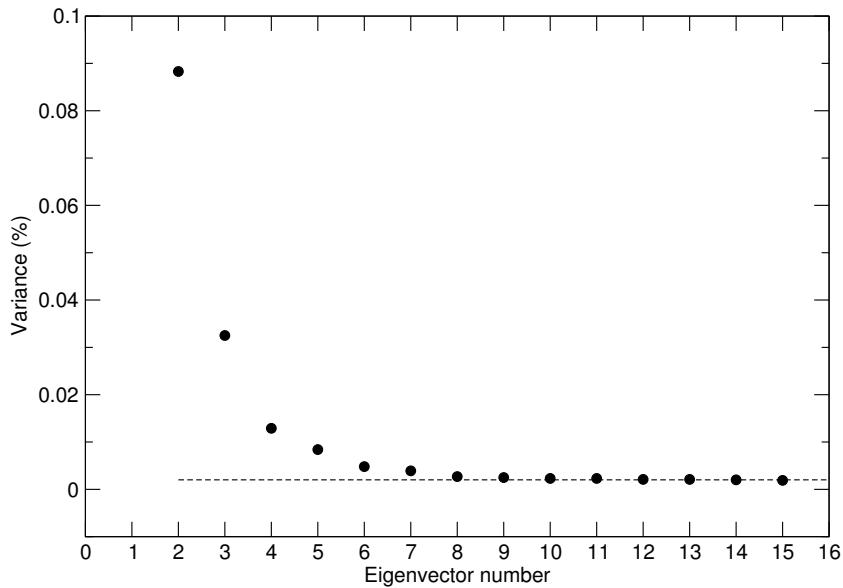


Figura 3.2: Scree test das primeiras 16 PCs do cubo de espectros da região central da galáxia NGC 4736. Retirado de Steiner et al. (2009, fig. 1).

sificação desses objetos como tipo 1 ou 2. O PCA, juntamente com a Tomografia PCA, fazem esse papel da subtração das populações estelares sem haver nenhuma parametrização.

Steiner et al estudaram os 8 primeiros autoespectros, escolhidos através de um *scree test* (3.2), no qual se verifica a variância de cada PC e toma-se as mais relevantes. O autoespectro com mais variância (no artigo tratado com E1) possui 99.74% da variância e reproduz comportamento do gás e da população estelar somados (Figura 3.3). O segundo contribui com 0.088% para a variância e tem um claro padrão de rotação, tanto nas linhas do autoespectro quanto na imagem da tomografia (Figura 3.4). É no terceiro (0.032% da variância) que mostra evidência clara de uma emissão larga de H α (Figura 3.5). Essa assinatura é uma evidência típica de AGNs de tipo 1.

3.3.2 Reflexao da luz de um AGN escondido em NGC 7097

!○j○! A FAZER!! !○j○! Ricci et al 2011. Figs 3 e 5. vais precisar da minha ajuda para entender esse assunto de reflexao ... le o paper e depois falamos

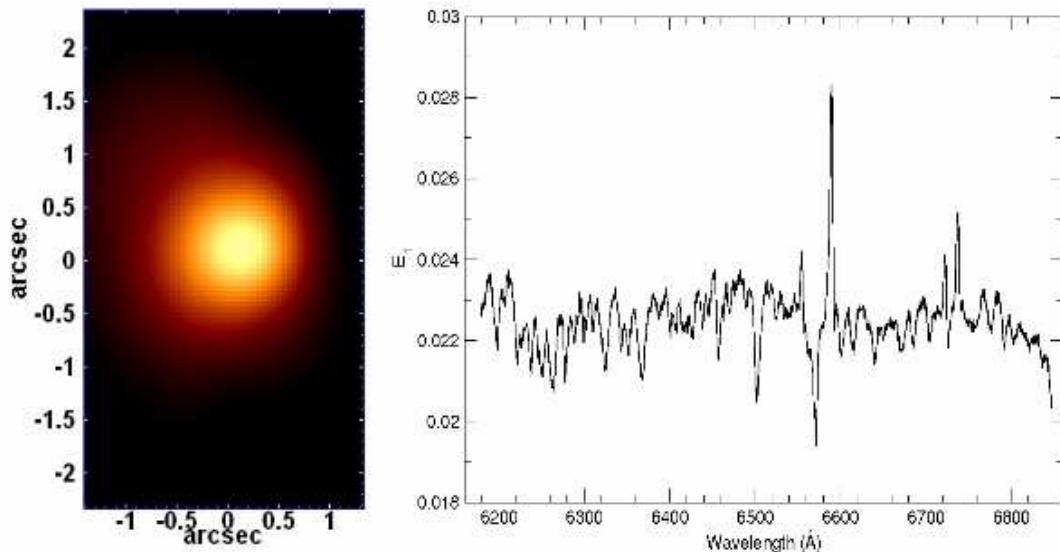


Figura 3.3: Autoespectro 1 e seu respectivo tomograma. Retirado de Steiner et al. (2009, fig. A1).

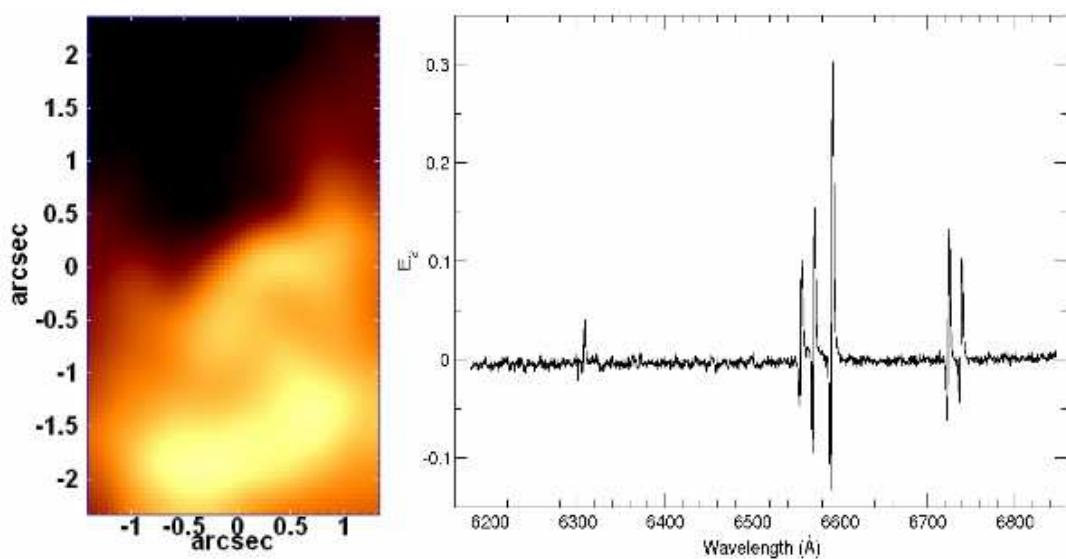


Figura 3.4: Autoespectro 2 e seu respectivo tomograma. Retirado de Steiner et al. (2009, fig. A2).

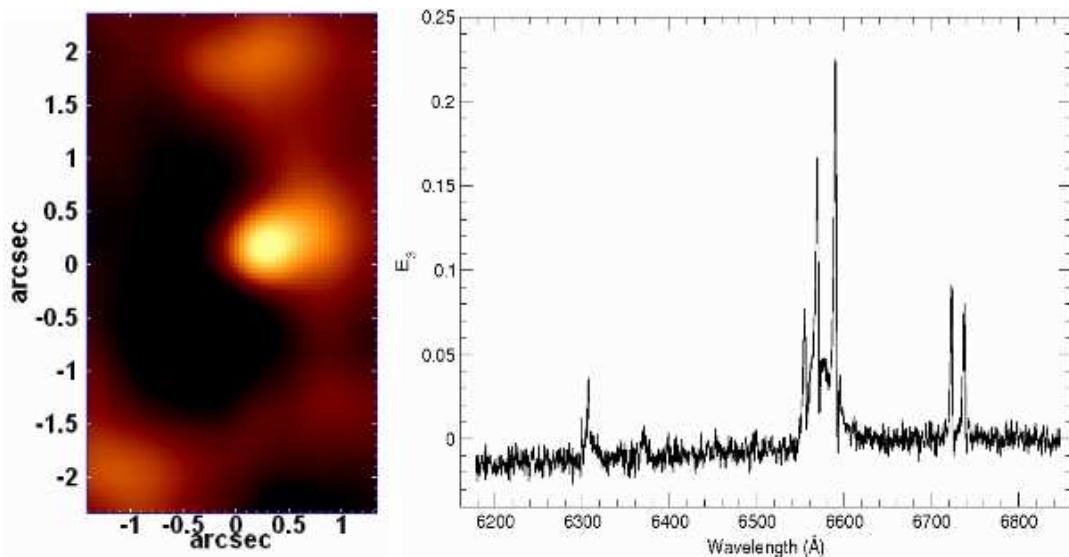


Figura 3.5: Autoespectro 2 e seu respectivo tomograma. Retirado de Steiner et al. (2009, fig. A3).

3.4 Tomografia PCA para dados CALIFA: diferenças com respeito a trabalhos anteriores

Os exemplos acima nos motivam a aplicar esta mesma técnica aos cubos de dados do CALIFA, e o resto desta dissertacão apresenta nossos expeimentos nesse sentido. Cumpre ressaltar, contudo, que o tipo de dados que analisaremos difere bastante dos dados analisados até agora com essa nova ferramenta.

Essas diferenças são tanto de caráter observacional, como metodológico e físico. Para começo de conversa, Steiner e colaboradores trabalham com dados do Gemini (8m) com ?? min de integracão, enquanto nossos dados vêm de integrações de tipicamente ?? min no telescópio de 3.5m do observatório de Calar Alto.

Além disso, os trabalhos de ?? e ?? aplicam técnicas de deconvolução das imagens com o algoritmo de Richardson-Lucy. Dado o pequeno FoV do IFU do Gemini, o ganho com tais técnicas é perceptível. Nao aplicaremos isso no nosso caso pq nosso FoV é muito maior, e não esperamos grandes ganhos em resolução espacial com a aplicação dessa técnica de deconvolução espacial.

Como dito na sec ???, a escala e resolução espacial de nossos dados difere brutalmente daquelas nos trabalhos acima resumidos: 3 x 5 arcsec e resolução de ?? arcsec no Gemini

contra ?? x ?? e resolução de 3.7 arcsec no CALIFA. Portanto, o FoV estudado por Steiner e colaboradores equivalente a aproximadamente 2 elementos de resolução do CALIFA! Isto de cara mostra que a ciência que podemos obter com Tomografia PCA + CALIFA não será a mesma que aquela estudada por Steiner et al.

Em resumo, apesar de inspirado diretamente pelos resultados de Steiner et al, este estudo difere muito dos trabalhos deles. Estas diferenças ficarão evidentes já a partir do próximo capítulo, onde apresentaremos nossos primeiros resultados e discutiremos algumas variações com respeito ao processamento de Steuer et al que decidimos fazer para melhor se adequar ao contexto de dados CALIFA.

Capítulo 4

Aplicação a cubos de dados CALIFA

Com os dados e ferramentas apresentados nos capítulos anteriores estamos prontos para aplicar a Tomografia PCA a galáxias do CALIFA. Como o PCA é uma técnica que calcula os eixos que, através da variância, melhor expandem a base de dados, é natural que qualquer pré-processamento que modifique os espectros de entrada acarrete mudanças na variância entre eles e portanto resulte num conjunto diferente de PCs. Durante nossas investigações fizemos uma série de testes com pré-processamentos e manipulações nos espectros (normalização, remoção de cinemática, janelas diferentes em comprimento de onda, PCA das linhas de emissão, logarítmico do fluxo, entre outros), verificando suas implicações no resultado. Neste Capítulo comentaremos alguns *insights* que tivemos nesse processo, e definiremos operações que julgamos adequadas a análise PCA de galáxias inteiras. Por fim faremos uma espécie de engenharia reversa, estudando correlações entre PCs e propriedades derivadas da síntese de populações estelares, o que nos auxilia na busca do sentido físico das PCs resultantes.

Neste Capítulo todas as análises serão feitas usando a galáxia NGC 2916 (CALIFA 277), uma espiral de tipo Sbc. A discussão detalhada desse caso particular serve de guia para os resultados para outras galáxias de nossa amostra, apresentados no Capítulo 5.

4.1 PCA dos espectros observados: Efeitos de amplitude

Como exemplo inicial apresentamos as cinco primeiras PCs e seus respectivos tomogramas do cubo de espectros observados da galáxia NGC 2916. Imagens SDSS e CALIFA dessa galáxia espiral podem ser vistas na Fig. 2.4, enquanto a Fig. 4.1 mostra alguns dos espectros

Figura 4.1: Alguns espectros do cubo de dados de NGC 2916. Os painéis à esquerda mostram os espectros originais (em erg/s/cm²/A), enquanto os painéis à direita mostra os mesmos espectros normalizados pelo fluxo mediano na janela de 5635 ± 45 Å.

!○j○! EDUARDO: escolher uns 5 a 10 espectros para esta fig!

CALIFA desse objeto.

A Figura 4.2) apresenta os resultados da tomografia PCA para esse cubo de dados. O tomograma é formado pelo “peso” da PC em cada zona. O primeiro autoespectro lembra o espectro médio (linha em cinza claro nos painéis à direita), embora muito mais azul. O segundo autoespectro se assemelha com um espectro de população muito jovem. Com o auxilio de seu tomograma vemos que para as zonas mais internas da galáxia (onde as populações sao mais velhas, como vimos nas Figs ??2.5 e 2.7??) o seu peso é negativo. De forma contrária, em regiões mais afastadas do núcleo, onde, para essa galáxia, parecem existir algumas regiões HII, vemos que seu peso é positivo. A terceira componente parece mostrar um padrão de rotação, ao mesmo tempo misturada com um fator de escala. Para que esse padrão de rotação ficasse visível tivemos que fazer um certo tratamento no tomograma, mas falaremos mais sobre isso no final dessa seção. Da quarta componente em diante fica mais complicado dizer o que cada uma representa. No final desse capítulo falaremos mais sobre como melhor interpretar as PCs fisicamente.

Da mesma forma, fazemos o PCA para o mesmo cubo, mas com os espectros normalizados da forma que comentamos em [\[FIXME\]](#) (Capítulo 2). Exemplos de espectros normalizados são mostrados nos painel direito da Fig. 4.1. A tomografia PCA desse cubo normalizado é mostrada na Figura 4.3. À primeira vista os resultados podem ser parecidos com o anterior deslocando uma componente acima, ou seja, o mesmo PCA mas sem a primeira PC. Mas essa é só uma primeira impressão. As propriedades físicas de uma galáxia geralmente possuem simetria radial ou axial, sendo assim extremamente correlacionadas no espaço. Matematicamente, a maior fonte de variância nos espectros é sua amplitude, aparecendo na forma de componente principal como visto na análise feita sem normalização. Essa componente de escala, quando não suprimida, correlaciona-se com as propriedades físicas da galáxia, portanto, deixa sua assinatura através de todas as PCs. Essa amplitude, que gera a primeira PC no caso sem normalização não adiciona nenhuma informação qualitativa à análise das populações estelares, funcionando apenas como um fator de escala, afetando as componentes seguintes. Esses vestígios de amplitude aparecem de forma mais evidente quando comparamos a PC3 no caso sem normalização com a PC2 no caso com normalização. Podemos ver um padrão

de rotação na PC2 da Figura 4.3. Com uma saturação na escala de cores no tomograma da PC3 da Figura 4.2 diminuímos o efeito do fator de escala ainda presente nessa componente, de modo que ficasse mais evidente o padrão de rotação. O espectro médio já é a informação necessária que precisaríamos para entender esse fator de escala nas amplitudes dos espectros, portanto a primeira componente para o caso sem normalização não trás informação adicional. Veremos a seguir que seu tomograma também não trás maiores informações.

Imagine uma galáxia composta inteiramente pela mesma população estelar, em repouso, distribuídas da mesma maneira no espaço. Ou seja, em qualquer ponto da galáxia o espectro é o mesmo. Um PCA nessa galáxia hipotética não identificaria nenhuma fonte de variância, pois todos os espectros são iguais ao espectro médio. Adicione então a essa galáxia uma função de densidade de massa estelar em função do raio (ou, equivalentemente, um perfil de brilho superficial), permitindo que de uma posição para outra a quantidade dessa determinada população se altera (mudando a intensidade da região). Uma componente nova irá surgir na sua análise PCA, mostrando que existe uma variância agora numa componente de escala (amplitude) nos espectros. Mas o que essa componente de escala nos diz sobre as propriedades da população estelar existente? Nada! Essa componente seria um desperdício de variância para uma análise de populações estelares de uma galáxia.

No caso do CALIFA, com um campo abrangendo praticamente toda a galáxia, esse efeito de amplitude se acentua muito devido ao brilho superficial mais intenso nas zonas centrais da galáxia em comparação com as mais afastadas, assim adicionando uma grande variância descartável entre as zonas. Descartável pois não trazem informação nova para a nossa análise. Essas diferenças em amplitude não nos dizem nada sobre as populações estelares. Comparemos agora a primeira PC do caso sem normalização (Figuras 4.2) e a imagem mais à esquerda na Figura 4.4 formada pelos fluxos para normalização por zona. Podemos notar que o primeiro autoespectro (e seu respectivo tomograma) no caso sem normalização mostra exatamente esse fator de escala. Seu tomograma mostra que ela é claramente um fator de escala (que pode ser considerado um fator de brilho, ou de amplitude, nos espectros).

Em suma, concluímos que é bem mais útil analisar espectros normalizados. Os trabalhos do grupo de J. Steiner não utilizaram esse esquema de normalização, mas isso é muito provavelmente devido ao pequeno campo coberto por seus dados. Nesse caso, as variações em amplitude são pequenas e não jogam um papel importante na análise. No caso do CALIFA, contudo, os efeitos de amplitude são muito maiores, e complicam a interpretação dos resultados da PCA. Por esse motivo, em nossas análises daqui para frente usaremos o cubo com os espectros normalizados.

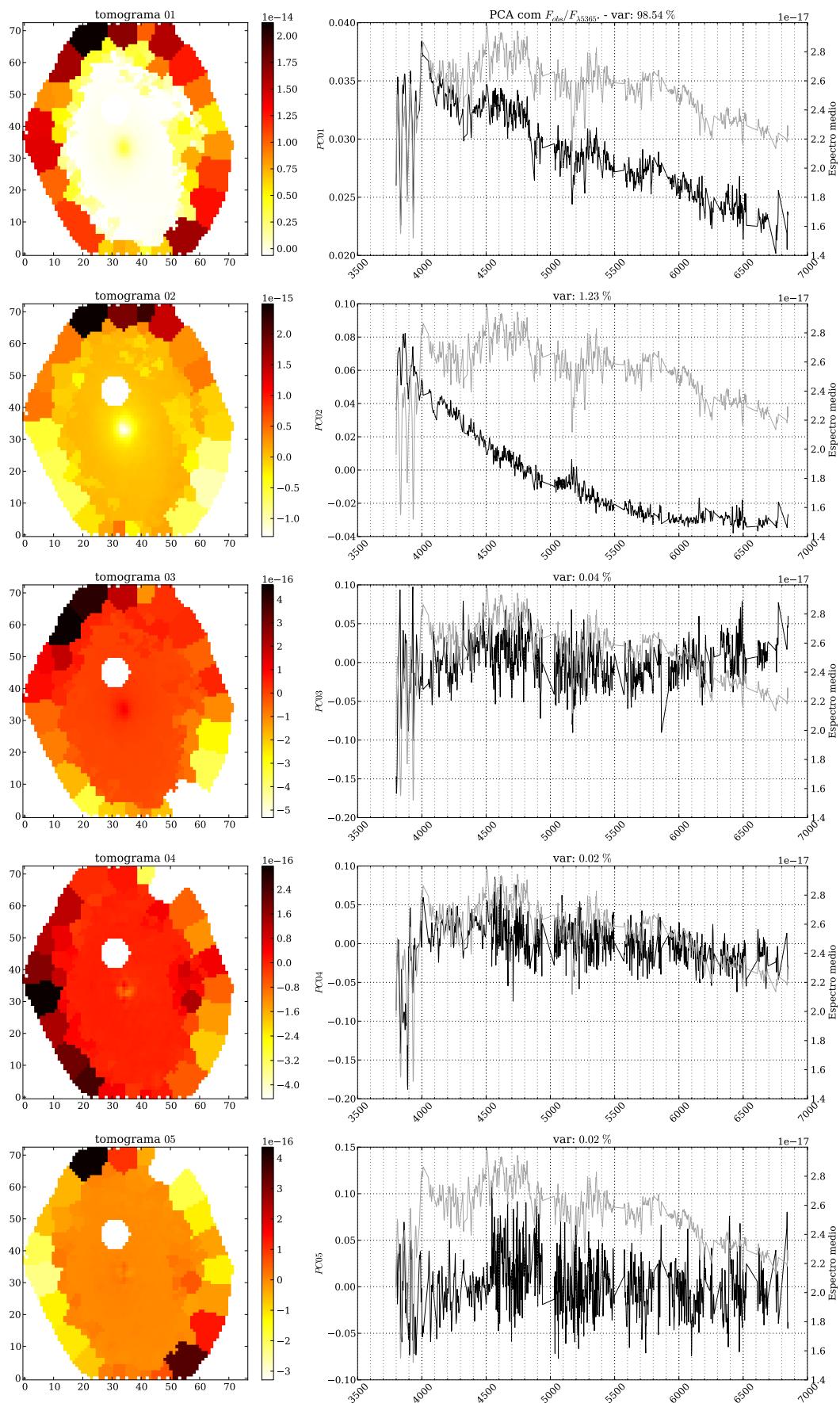


Figura 4.2: Cinco primeiras PCs (e seus respectivos tomogramas) resultantes da Tomografia PCA aplicado aos espectros sem normalização da galáxia NGC 2916

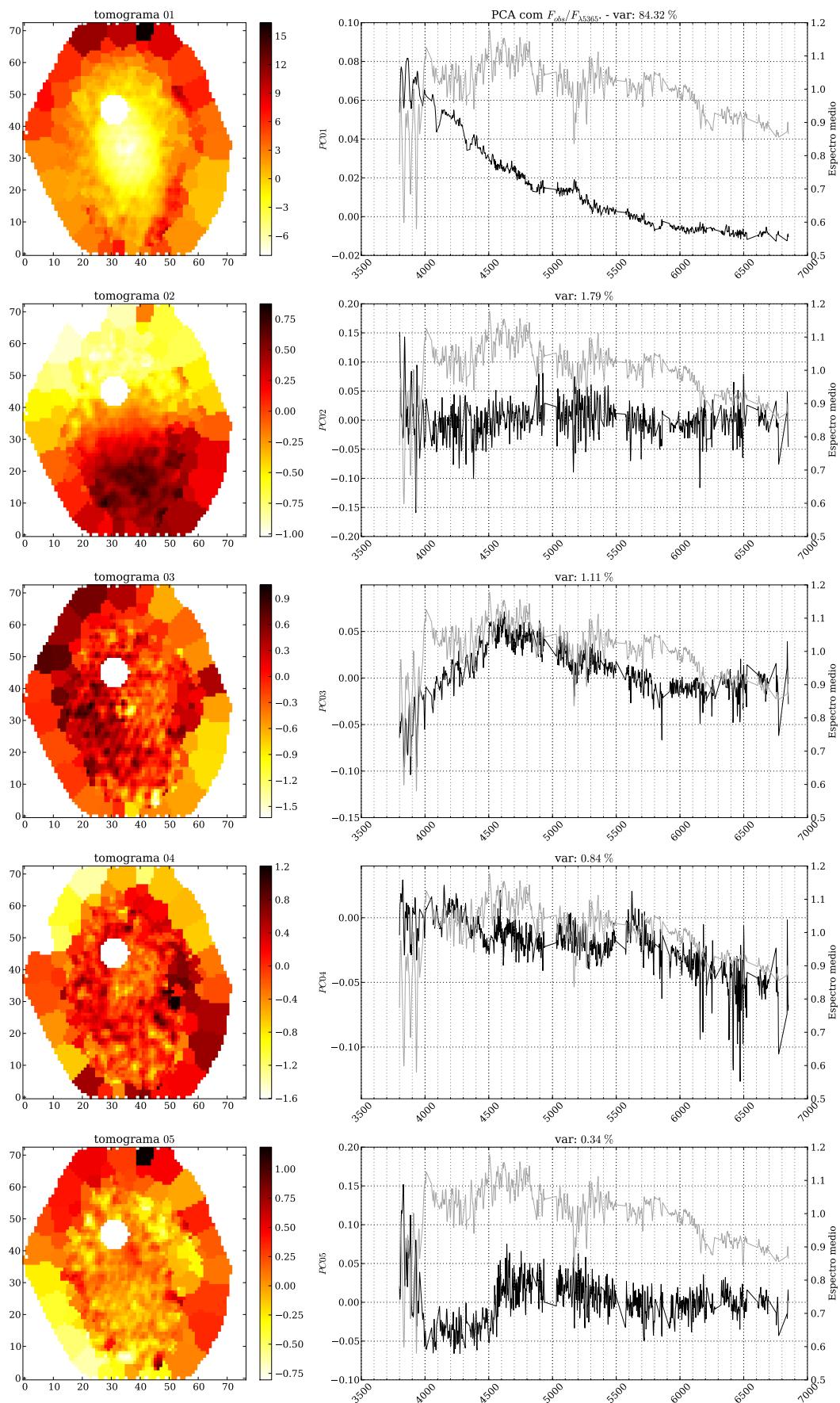


Figura 4.3: Cinco primeiras PCs (e seus respectivos tomogramas) resultantes da Tomografia PCA aplicado aos espectros com normalização da galáxia NGC 2916

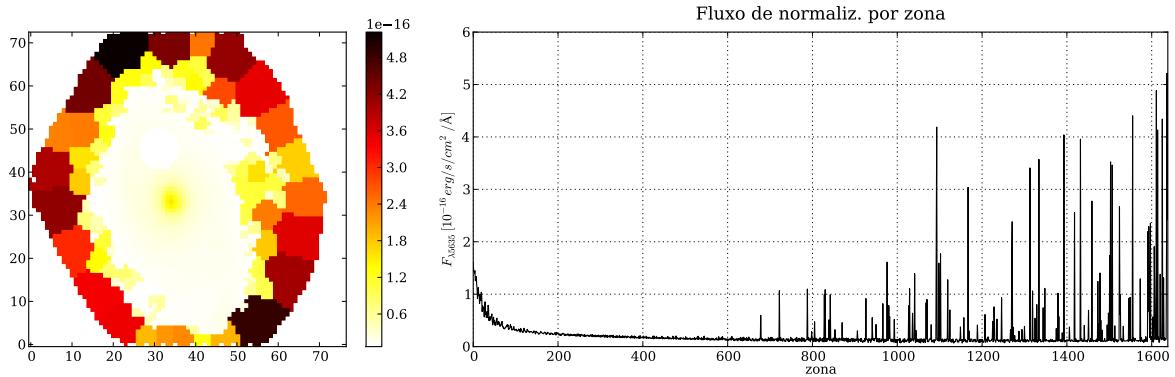


Figura 4.4: Fluxo usado para a normalização de cada espectro, mostrado tanto na forma de imagem (à esquerda) como em função do número da zona (direita).

4.2 PCA dos espectros sintéticos

Na seção anterior realizamos o PCA no cubo de espectros observados de CALIFA 277. Com o resultado da síntese de populações estelares já organizado com o pipeline PyCASSO podemos aplicar a mesma técnica aos espectros sintéticos gerados pelo STARLIGHT.

A grande diferença é que nos espectros sintéticos estão contidas apenas as informações sobre populações estelares¹. Como os espectros sintéticos não possuem as assinaturas dos equipamentos observacionais, dos processos para subtração do céu, ruídos e afins, quando analisadas pela técnica de PCA esperamos que as informações se condense em menos PCs. Comparando os dois *scree tests* na Figura 4.5 vemos que para o caso com os espectros sintéticos a curva converge mais rapidamente ao zero de variância, mostrando que, como esperado, temos as informações mais compactadas nas primeiras PCs quando comparadas ao caso com os espectros observados. Observando o caso sem normalização, plotado no gráfico com linha pontilhada, vemos que o efeito causado pelo fator de escala (PC1) diminui a “importância” (i.e., variância) das demais PCs.

As cinco primeiras PCs e seus tomogramas provenientes do cubo com os fluxos sintéticos normalizados da galáxia NGC 2916 aparecem na Figura 4.6.

Comparando com seu correspondente observational na Fig. 4.3, vemos que os resultados para os espectros sintéticos parecem ser mais “limpos”, pois não existem ruídos. Como são espectros gerados a partir de uma base teórica para diferentes idades e metalicidades de

¹Efeitos de extinção/avermelhamento e cinemática também são incluídos no processo de síntese ((Cid Fernandes et al. 2005)

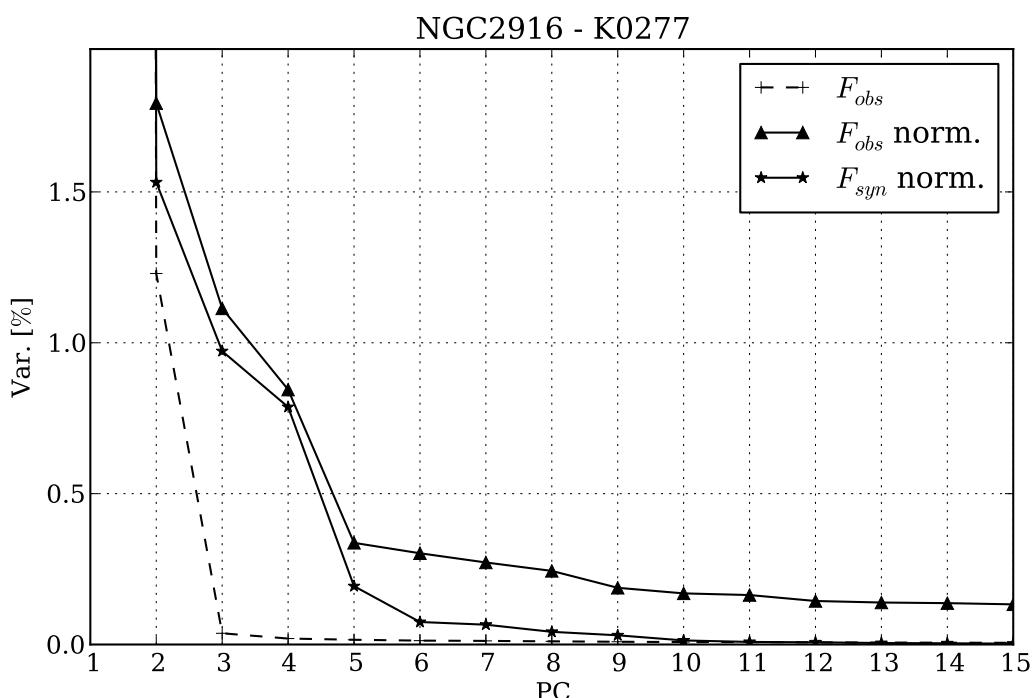


Figura 4.5: Scree test para 3 análises PCA da galáxia NGC 2916 (CALIFA 277). Com marcações de triângulos vemos as PCs resultantes do PCA com os espectros observados normalizados. As variâncias das PCs marcadas com estrela representam o PCA com os espectros sintéticos normalizados. Para comparação plotamos as PCs do caso sem normalização usando linha pontilhada.

populações estelares, não possuem nenhuma assinatura instrumental. Por esse fato acabamos descobrindo uma assinatura presente em quase todas as componentes geradas pelo PCA usando os dados observados. Como comentamos no Capítulo 2.1 utilizamos o cubo de dados COMBO, gerado a partir da união do V500 com o V1200. Como os espectros V1200 possuem espectros com maior resolução do que os do V500 (V1200 - FWHM $\sim 2.3 \text{ \AA}$; V500 - FWHM $\sim 6 \text{ \AA}$), o processo de criação do COMBO deixa vestígios. Para essa galáxia, a junção entre os dois cubos (V500 e V1200) para a formação do COMBO acontece exatamente nesse intervalo de comprimento de onda ($\sim 4550 \text{ \AA}$). O quinto autoespectro da Figura 4.3 mostra um degrau entre 4500 e 4600 \AA ao qual parece mostrar essa diferença de comportamento entre as duas versões originais antes da formação do COMBO. Já nas componentes sintéticas² não se vê essa mudança de comportamento no autoespectro.

Assim como na análise dos espectros observados normalizados, os efeitos da cinemática se fazem notar já na segunda PC dos cubos sintéticos, indicando que são fonte importante de variância. Consideramos, porém, que essa é uma variância "descartável". Usando novamente a ideia da galáxia hipotética com apenas uma população estelar, imagine agora que elas estão distribuídas uniformemente, mas estão em rotação com a galáxia. Como anteriormente, o espectro de todos os pixels será igual, mas agora terá deslocamentos em λ . Esses efeitos cinemáticos não estão nos trazendo informação alguma para o estudo das populações estelares. Causam um grande desperdício de variâncias, sempre aparecendo nas primeiras PCs. Uma manipulação através do vetor de população³ criado pela síntese, juntamente com os espectros da base, pode nos ajudar a criar os espectros sintéticos sem nenhuma correção por cinemática e poeira de maneira que se execute o PCA sem o desperdício de variância dessas componentes cinemáticas, mas este experimento ainda não foi realizado. Também é importante lembrar que existem métodos mais eficazes e direcionados para a determinação de tais propriedades cinemáticas. Nas galáxias presentes no CALIFA não poderia ser diferente, portanto os espectros aparecem com linhas deslocadas para o azul (*blue-shifted*) ou para o vermelho (*red-shifted*) dependendo da velocidade de rotação projetada. A dispersão de velocidades em cada ponto da galáxia também pode causar alargamento ou estreitamento das linhas. **!○j○! CONFUSO....**

²Componentes geradas pelo PCA nos cubos de espectros sintéticos.

³O vetor de populações diz o quanto de cada população estelar da base entra na receita para construir o espectro sintético.

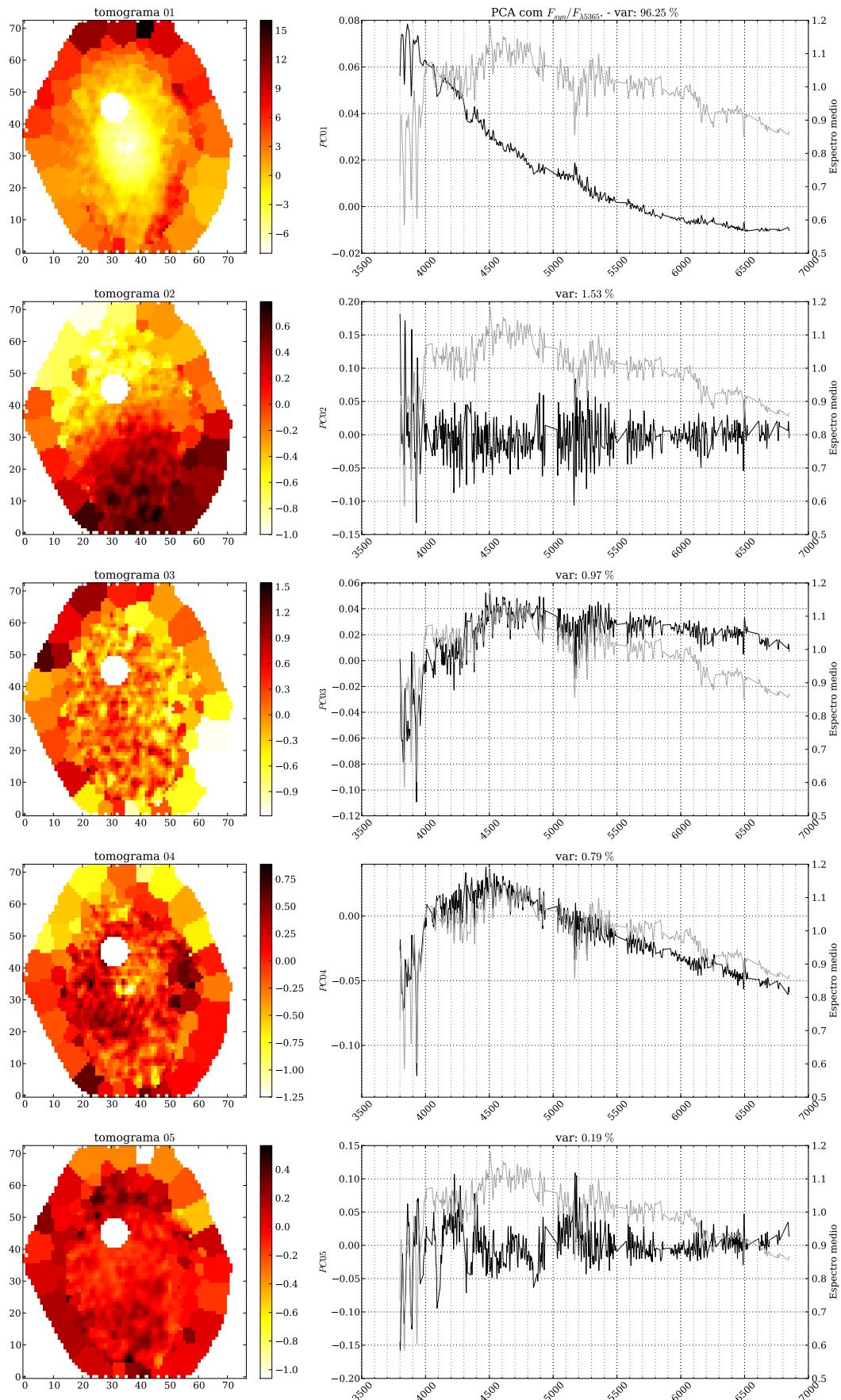


Figura 4.6: Cinco primeiras PCs (e seus respectivos tomogramas) resultantes da Tomografia PCA aplicado aos espectros sintéticos normalizados da galáxia NGC 2916

4.3 PCA logarítmica

Um outro experimento que realizamos foi calcular o PCA sobre os logarítmos dos espectros. Essa variante da analise pode ser vista como uma mera curiosidade matematica, uma experienca para ver o que se obtém. No entanto, ela foi inspirada em um motivo físico: É razoavel esperar que utilizando o log possamos ressaltar efeitos devidos a gradientes de poeira, já que a extinção que atua de forma multiplicativa sobre os espectros.

Para entender melhor voltemos ao caso hipotetico de uma galaxia na qual as populações estelares são iguais em todas partes e na qual as estrelas estão todas paradas, de modo que não há deslocamentos nem modulações em λ devido a efeitos de cinematica. Neste caso, as variações de $\mathbf{F}_{z\lambda}$ entre as diferentes zonas vem tanto do efeito de amplitude (brilho superficial) já estudado como de diferenças na profundidade óptica da poeira $\tau_{z\lambda}$. O efeito global de amplitude pode ser removido com nosso esquema de normalização, mas as variações de $\tau_{z\lambda}$ de zona a zona afetam a forma do espectro observado.

O efeito da poeira é atenuar um fluxo intrínseco F^i por um fator $e^{-\tau}$, produzindo um fluxo observado (transmitido) F^o . Escrevendo $F^o = F^i e^{-\tau}$ para um certo λ e normalizando-o pelo comprimento de onda de normalização λ_N obtemos

$$\frac{\mathbf{F}_{z\lambda}^o}{\mathbf{F}_{z\lambda_N}^o} = \frac{\mathbf{F}_{z\lambda}^i}{\mathbf{F}_{z\lambda_N}^i} \frac{e^{-\tau_{z\lambda}}}{e^{-\tau_{z\lambda_N}}} \quad (4.1)$$

Denotando os espectros normalizados por O_λ e assumindo que a lei de extinção é a mesma em todas zonas

$$\mathbf{O}_{z\lambda}^o = \mathbf{O}_{z\lambda}^i e^{-\tau_{zV}(q_\lambda - q_{\lambda_N})} \quad (4.2)$$

onde $q_\lambda \equiv \tau_\lambda / \tau_V$ regula o grau de avermelhamento. Tirando o ln

$$\ln \mathbf{O}_{z\lambda}^o = \ln \mathbf{O}_{z\lambda}^i - \tau_{zV}(q_\lambda - q_{\lambda_N}) \quad (4.3)$$

Se $\mathbf{O}_{z\lambda}^i$ é igual para todas zonas, como é o caso em nossa galáxia hipotética, fica claro que toda variancia em $\ln \mathbf{O}_{z\lambda}^o$ estará associada a variancia de τ_{zV} de zona-a-zona, e a tomografia PCA mapeará essa variancia. Sem as operações de normalização e logaritmo, esse efeito apareceria misturado com outros.

CALIFA 277, a galáxia que temos usado como exemplo ao longo desse capítulo, contém pouca poeira, não sendo portanto um bom caso para aplicação dessa variante. Exemplificaremos os resultados da PCA logarítmica no capítulo 5 ao estudarmos os sistemas em fusão (*mergers*) Arp 220 e NGC ????, ricos em poeira.

4.4 Comparando as PCs com o *STARLIGHT*: engenharia reversa

Descobrir o sentido físico de cada PC não é tarefa fácil. Como comentamos anteriormente, o PCA te dá a resposta, mas você geralmente não sabe qual a pergunta. Uma forma de buscar sentido físico nas componentes é analisar as correlações com propriedades físicas da galáxia.

Com o resultado da síntese de populações estelares obtido pelo *STARLIGHT* e organizado pelo PyCASSO fica simples correlacionarmos os dados na base gerada pelo PCA. Nas Figuras 4.7 e 4.8 vemos as correlações entre algumas propriedades físicas ($\langle \log t \rangle$, $\log \langle Z \rangle$, A_V , v_* , σ_*) e o peso de cada PC por zona (tomograma versus propriedade física) para CALIFA 277. O número presente em cada gráfico é o coeficiente de correlação por *rank* de Spearman. Escolhemos o coeficiente de Spearman pois este é aparamétrico e, ao contrário do coeficiente de Pearson, não pressupõe nenhuma relação linear. Em cada coluna temos uma quantidade física e na última coluna temos o autovetor representado por aqueles pontos.

Tanto para os dados observados quanto sintéticos vemos que a primeira PC representa basicamente um fator de idade. Possui correlação também com a metalicidade. A segunda PC também em ambos os casos representa uma forte correlação com a velocidade estelar, refletindo a curva de rotação da galáxia. A terceira PC com os dados observados parece ser um bom sensor para o padrão global de extinção (A_V). Essa PC parece se dissolver em duas no caso do espectro sintético (PC3 e 4). As PC4 e PC5 no caso observado não parecem ter correlação com nenhum desses parâmetros físicos comparados, salvo pequenas correlações com A_V e v_* na PC5. Essa última, para o caso sintético, parece estar correlacionada com A_V também, mas de forma mais fraca. Também possui um padrão de correlação com v_* . Por fim, a PC6 no caso observado parece ter uma mistura de idade, metalicidade, extinção e com a

disperção de velocidades. Para o caso sintético existe uma correlação com a metalicidade e a dispersão de velocidades. Um padrão de rotação também aparece mas com um coeficiente de spearmann baixo.

Outra forma de se observar essas correlações é graficando os dados nos eixos das PCs, colorindo os pontos por determinado parâmetro físico. No Apêndice estarão todos os gráficos com todos os parâmetros físicos. Como exemplo aqui deixamos as Figuras 4.9 e 4.10 que mostram as PCs para o caso observado e sintético, coloridos pelo A_V .

A Figura 4.11 mostra também o efeito que a falta de normalização faz com o resultado do PCA. Esse fator de amplitude que não é removido pela normalização se concentra principalmente na primeira PC, mas ainda deixa vestígios se misturando com as outras PCs fazendo com que haja correlação entre cada PCs e quase todos os parâmetros físicos. Por mais esse motivo, deixaremos a PCA dos espectros não normalizados de lado no restante desse trabalho.

.....

DAQUI PRA BAIXO AINDA É O ESQUELETO !obj! A Filosofia desse capítulo é aprender/testar como operar o PCA para que ele reflita isso ou aquilo... descrevemos uma série de experimentos nesse sentido.

Preprocessamentos e diferentes tipos de PCAs com ou sem linhas, diferentes faixas especiais, com dados normalizados ou não. (importante!!!)

Vamos nos limitar a no-emission lines analysis, descrever a máscara de linhas de emissão, etc. Isso para facilitar a coisa e pq queremos correlar o resultado do PCA com os dados do Starlight (PyCASSO).

Simulações para ajudar a decifrar os resultados, população jovem + velha + modelo de distribuição espacial - ver efeitos de estratégias de preprocessamento.

Correlacionando os resultados do PCA com as propriedades do starlight (tipo de engenharia reversa)

Linhos telúricas - remover ou não... bad pixels... mascarar ou não linhas de emissão

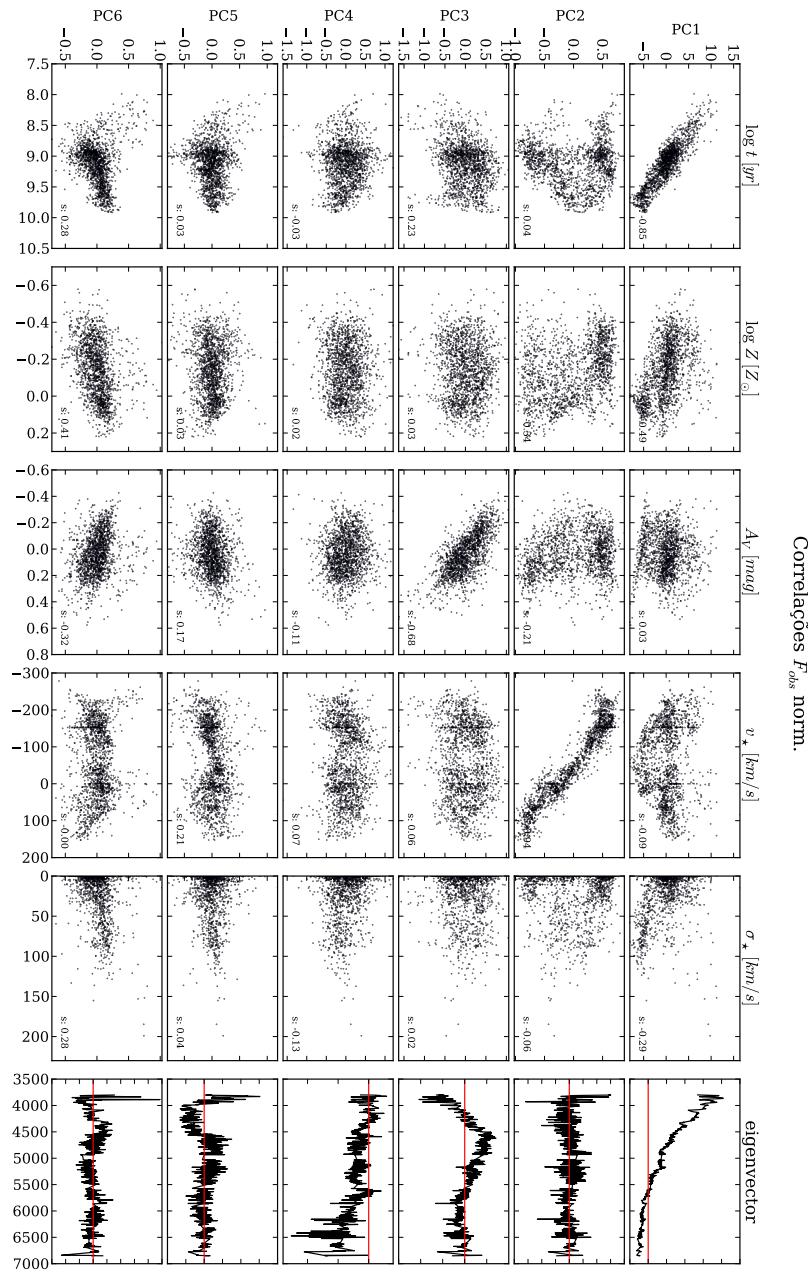


Figura 4.7: Correlações entre os pesos por zona das seis primeiras PCs do PCA feito para o cubo com os dados observados normalizados e cinco parâmetros físicos. Pela ordem de colunas da esquerda para direita temos $\log t$, $\log Z/Z_\odot$, A_V , v_* , σ_* . Na última coluna temos o autoespectro para ajudar na visualização. A linha em vermelho no gráfico do autoespectro serve para identificar o zero. O número dentro de cada gráfico é o coeficiente de correlação de Spearman.

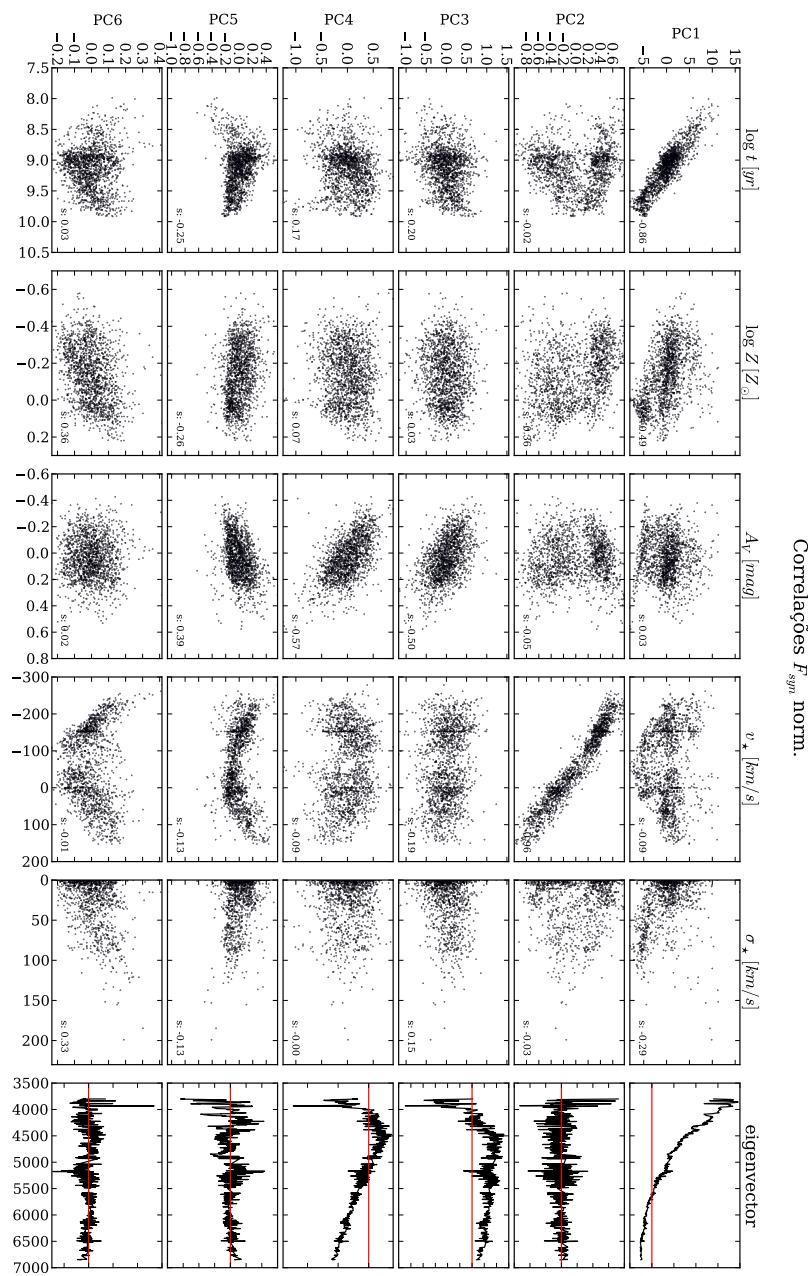


Figura 4.8: Correlações entre os pesos por zona das seis primeiras PCs do PCA feito para o cubo com os dados sintéticos normalizados e cinco parâmetros físicos.^q Pela ordem de colunas da esquerda para direita temos $\log t$, $\log Z/Z_\odot$, A_V , v_* , σ_* . Na última coluna temos o autoespectro para ajudar na visualização. A linha em vermelho no gráfico do autoespectro serve para identificar o zero. O número dentro de cada gráfico é o coeficiente de correlação de Spearman.

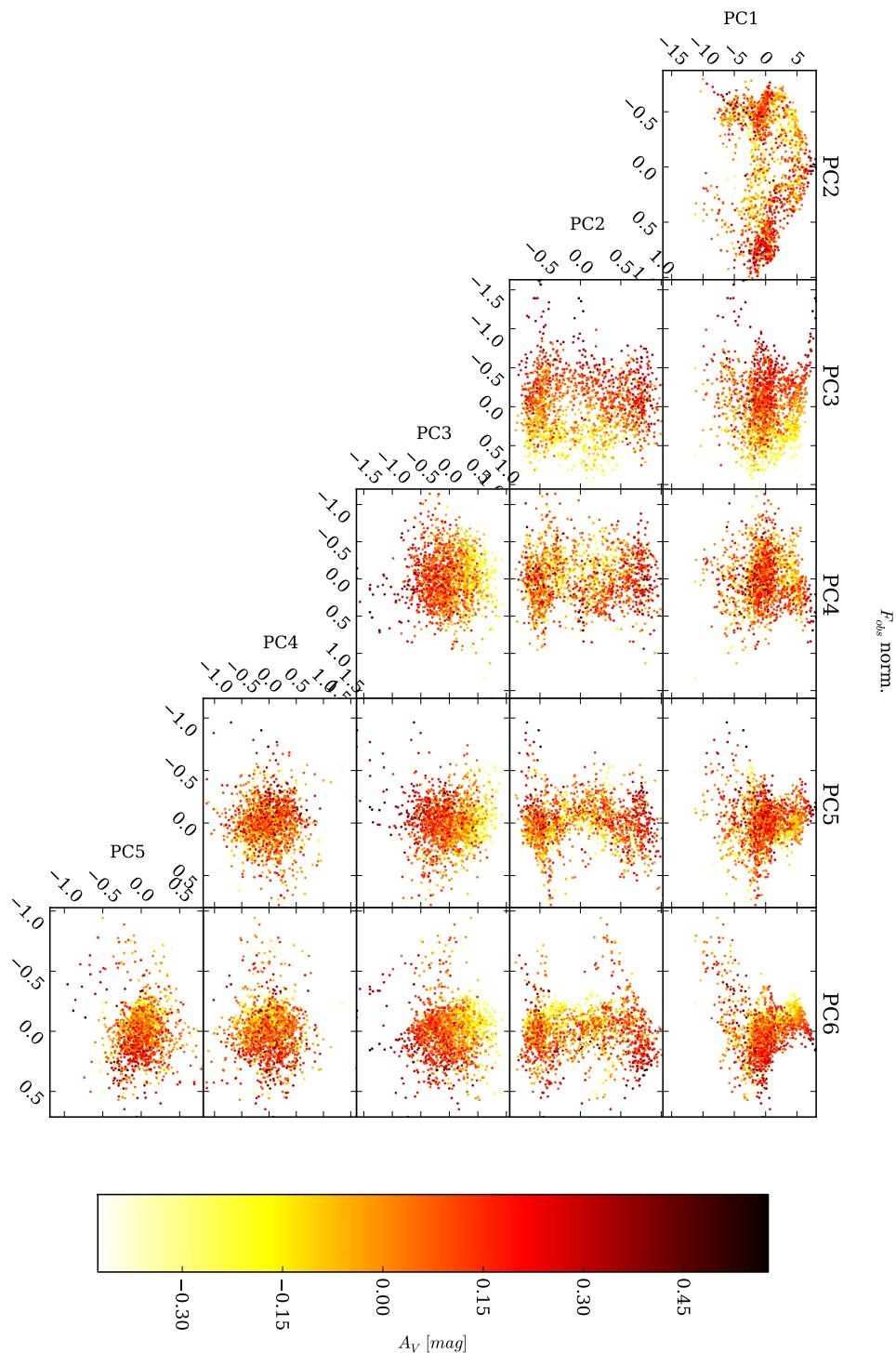


Figura 4.9: Dados das zonas no espaço das PCs (com espectros observados) coloridos pela extinção (A_V).

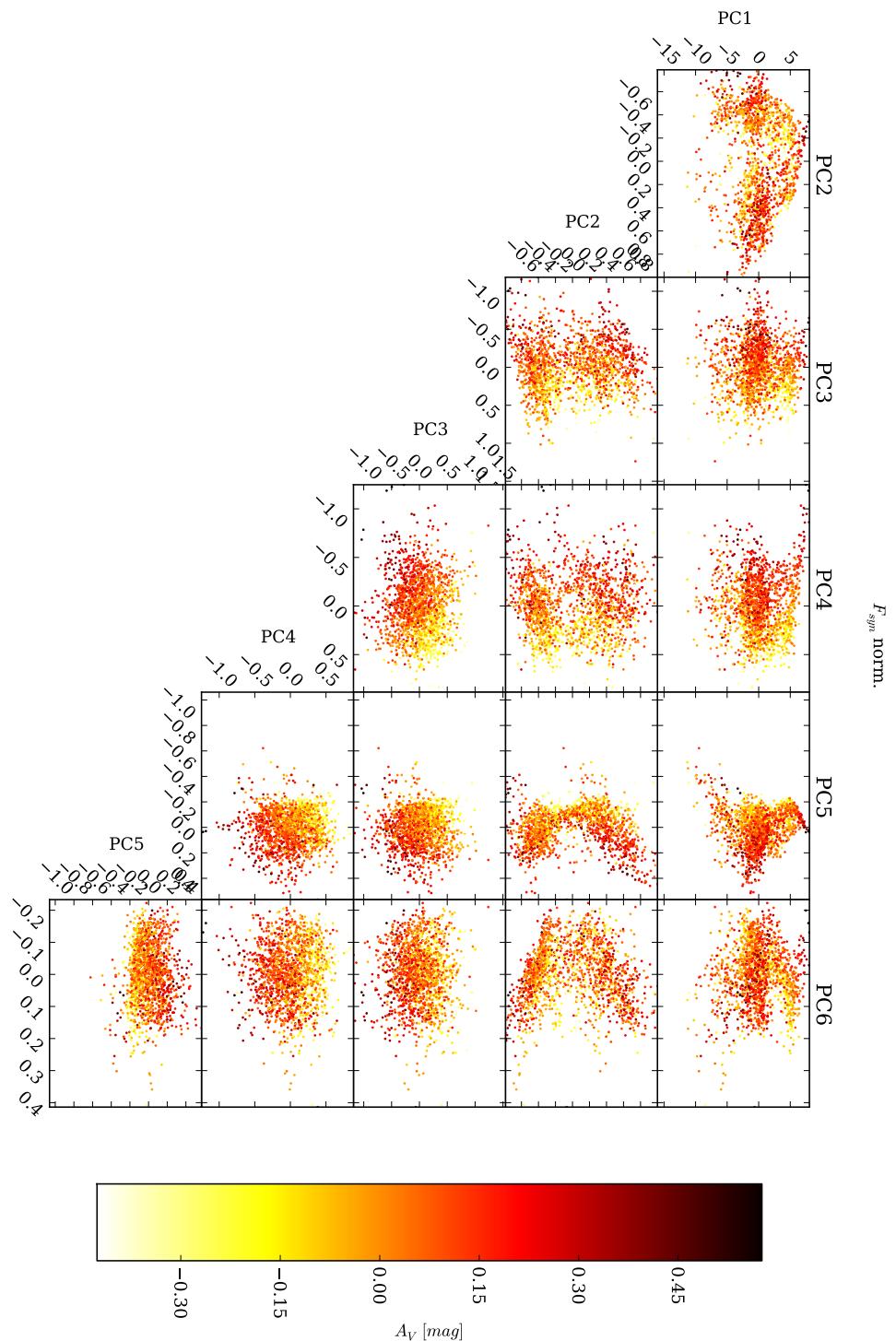


Figura 4.10: Dados das zonas no espaço das PCs (com espectros sintéticos) coloridos pela extinção (A_V).

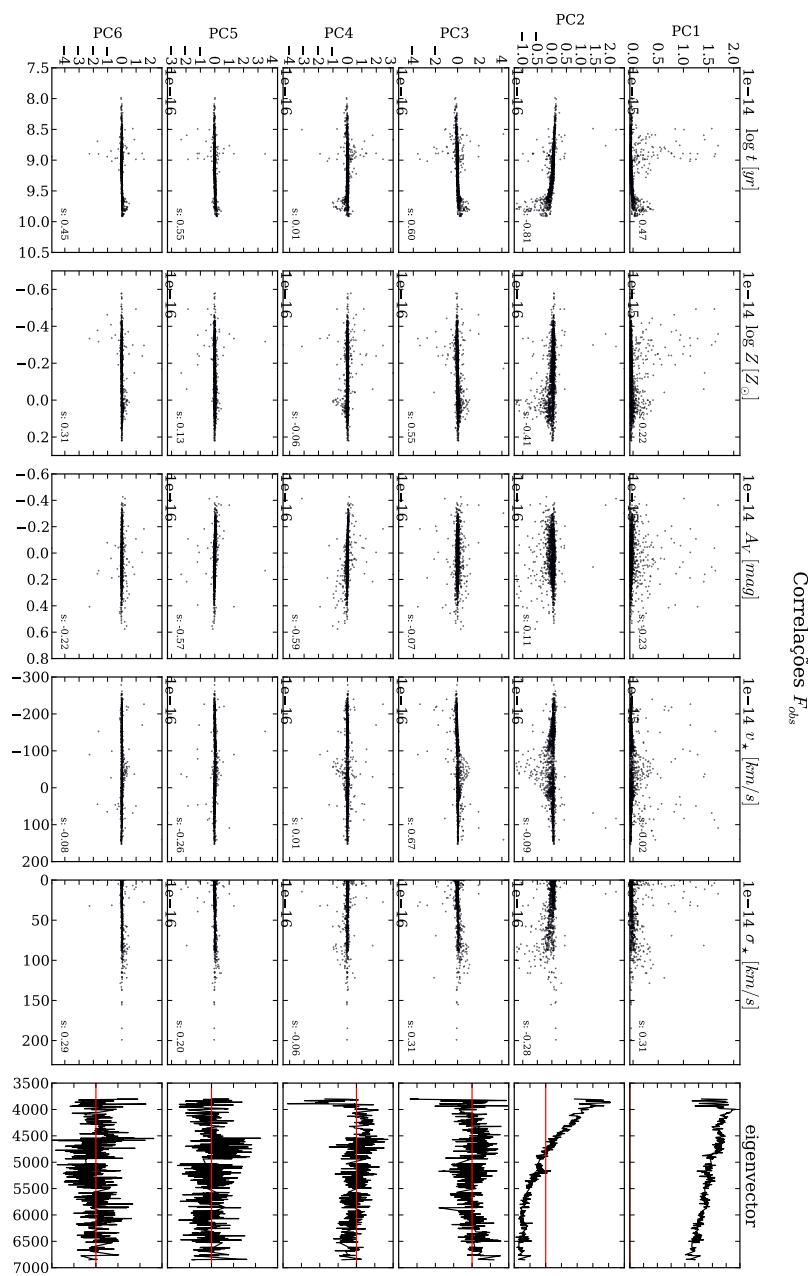


Figura 4.11: Correlações entre os pesos por zona das seis primeiras PCs do PCA feito para o cubo com os dados observados e cinco parâmetros físicos. Pela ordem de colunas da esquerda para direita temos $\log t$, $\log Z/Z_\odot$, A_V , v_\star , σ_\star . Na última coluna temos o autoespectro para ajudar na visualização. A linha em vermelho no gráfico do autoespectro serve para identificar o zero.

Capítulo 5

Aplicando a Tomografia PCA em 8 galáxias do CALIFA

Com a possibilidade da aplicação de técnicas estatísticas entre as populações estelares distribuídas em uma mesma galáxia, abarcando, em campo de visão, toda galáxia, e com toda a discussão sobre as diferentes configurações dos dados no capítulo anterior, temos munição para a primeira exploração científica usando Tomografia PCA em galáxias do CALIFA. Nesse capítulo aplicaremos a Tomografia PCA em algumas galáxias de maneira exploratória, para que possamos identificar determinados padrões de variâncias nos dados. Por fim usaremos de nossa engenharia reversa, correlacionando as PCs com determinadas propriedades físicas de populações estelares obtidas pela síntese.

!♂! Em alguns casos podemos encontrar até problemas que passaram pelo processo de redução, marcação dos píxels problemáticos, subtração de poeira, em determinado espectro ou cubo.

5.1 Apresentação das galáxias

Dentre as galáxias já observadas pelo projeto CALIFA, escolhemos quatro espirais (NGC 0001, NGC 2916, NGC 0776 e NGC4210), duas galáxias *early-type* (NGC 1167 e NGC 6515) e dois *mergers* (NGC 2623 e ARP 220). Omitimos nesse capítulo as imagens referentes a análise da galáxia NGC 2916 pois já se encontram no capítulo anterior. **[FIXME]** **Imagen de apresentação da K0277 no cap anterior.** A sessão com mergers colocamos para pura “diversão”. Esses lindos objetos, produtos de algum grande encontro (ou alguns) de galáxias, são

extremamente difíceis de serem estudados e altamente complexos. Todas as galáxias aqui estudadas estão listadas na Tabela 5.1 na ordem que aparecerão neste capítulo.

Tabela 5.1: Tabela contendo a relação de galáxias que vamos estudar nesse capítulo, juntamente com sua classificação morfológica (*Hubble Type*), massa em estrelas, *redshift* e número de zonas.

Nome da galáxia	CALIFA ID	<i>Hubble Type</i>	log Mass [M_{\odot}]	redshift	N_z
NGC 0001	K0008	Sbc	11.00	0.01515	1132
NGC 2916	K0277	Sbc	10.83	0.01244	1638
NGC 0776	K0073	Sb	11.19	0.01640	1733
NGC 4210	K0518	Sb	10.49	0.00906	1938
NGC 1167	K0119	S0	11.47	0.01645	1879
NGC 6515	K0864	E3	11.42	0.02285	887
NGC 2623	K0213	Scd	10.74	0.01847	561
ARP 220	K0802	Sd	11.15	0.01814	1157

5.2 Galáxias espirais

5.2.1 NGC 0001 - CALIFA 8

No capítulo passado escolhemos como exemplo a galáxia NGC 2916 pelo fato que é uma das galáxias, do projeto CALIFA, mais estudadas pelo nosso grupo até agora (Cid Fernandes et al. 2013, 2014). Aqui começaremos falando sobre a NGC 0001 (CALIFA 8).

Podemos ver na Figura 5.1 a imagem obtida com o SDSS para a galáxia, o fluxo em 5635 Å, por zona, usado para a normalização, assim como imagens geradas a partir das propriedades físicas resolvidas zona a zona pela síntese. Idade média estelar, metalicidade, avermelhamento, velocidade estelar e dispersão de velocidades são as propriedades físicas.

Os *scree tests* das análises PCA para esta galáxia aparecem na 5.2 e mostram o mesmo padrão discutido em ??, onde o resultado do PCA para F_{syn} normalizado converge mais rápido ao zero de variância. É interessante notar também que apesar da quinta componente ainda ter 2% da variância para o caso F_{obs} normalizado, através da nossa análise dos tomogramas as componentes daí pra frente não parecem ter mais informações “decifráveis”.

Como a NGC 2916, a primeira PCs, e seu respectivo tomograma (Figuras 5.3 e 5.4) parece ser relacionada a uma população muito azul, tanto para o PCA com os dados observados, quanto sintéticos, correlacionando geralmente com a idade. Mas, vemos nas correlações (Figuras 5.5 e 5.6), que a primeira componente, para essa galáxia, se correlaciona praticamente com todas as quantidades avaliadas, salvo v_* . A idade média aparece mais definida junto com o A_V na PC4 para o caso observado. Já no caso sintético ela aparece melhor na PC6, juntamente com metalicidade e A_V , como índices maiores de correlação. Observado a imagem da velocidade estelar (Figura ??, segunda imagem na última linha) podemos ver que essa galáxia não possui um padrão muito claro de rotação, mas na análise das correlações para os dados sintéticos vemos que a PC4 se correlaciona de maneira formidável com v_* ($s = 0.88$). Para o caso observado vemos que a rotação aparece, mesmo que com uma correlação mais fraca que no caso sintético, na PC3.

5.2.2 NGC 0776 - CALIFA 73

A galáxia NGC 0776, apresentada na Figura 5.7, é uma espiral que possui uma diferente v_* para cada zona de uma forma mais suave e quase constante, diferentemente das duas já apresentadas até agora (NGC 2916 e NGC 0001).

A análise da variância percentual de cada componente através do *scree test* (Figura 5.8) segue o mesmo padrão de todas apresentadas até agora, sempre com a convergência ao zero de variância seguindo a ordem primeiro F_{obs} , depois F_{syn} normalizado e por último F_{obs} normalizado. Veremos que todos os *scree tests* avaliados seguem esse mesmo padrão assintótico, independentemente do tipo morfológico da galáxia.

Ambas análises para diferentes configurações resultam na primeira PC muito semelhante (Figuras 5.9 e 5.10) que correlaciona de maneira forte com a idade média (Figuras 5.11 e 5.12). A segunda componente para o caso sintético parece carregar informação da série de Balmer (também vista na PC4) e correlacionar com A_V . Já para o caso observado possui melhor correlação com a velocidade estelar. A informação sobre A_V parece estar melhor representada pela PC3 para o caso observado, diferentemente do caso sintético que parece conter essa informação de maneira um pouco mais clara na PC4. Vale a pena ressaltar que apesar dessa galáxia não ter um campo de velocidades muito distribuído, a PC5 para o caso sintético parece ser uma boa sonda dessa propriedade.

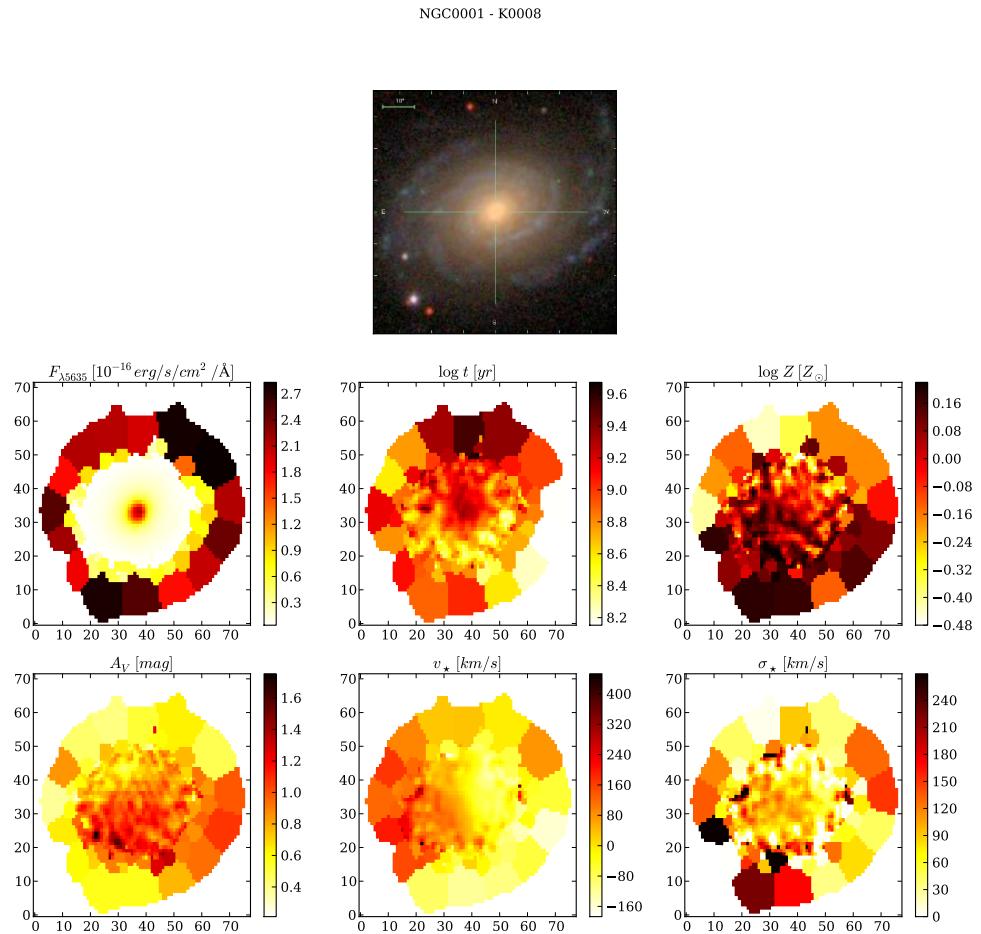


Figura 5.1: Propriedades físicas da galáxia NGC 0001. Na primeira linha a imagem do SDSS da galáxia. Na primeira imagem da segunda linha temos o valor do fluxo observado em 5635 \AA , usado para a normalização dos espectros de cada zona. Da segunda imagem da segunda linha em diante temos as propriedades físicas que serão correlacionadas com as PCs. São elas: $\log t$, $\log Z$, A_V , v_{\star} , σ_{\star} .

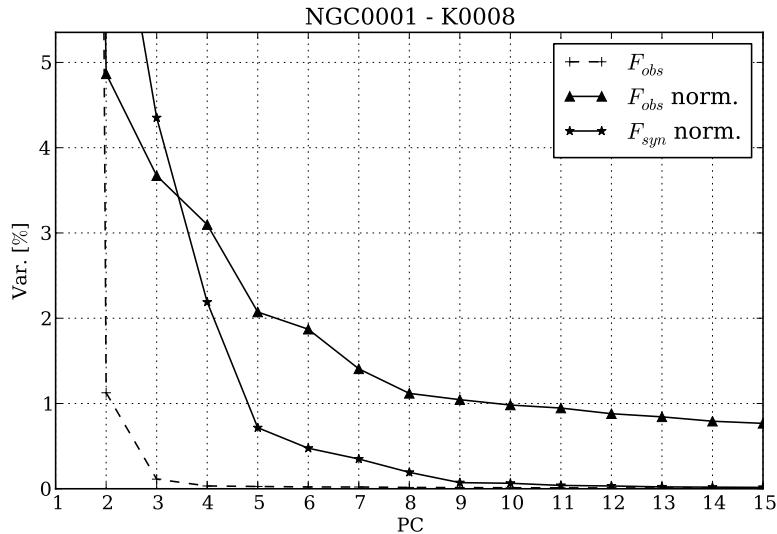


Figura 5.2: Scree test para 3 análises PCA da galáxia NGC 0001 (CALIFA 8). Com marcações de triângulos vemos as PCs resultantes do PCA com os espectros observados normalizados. As variâncias das PCs marcadas com estrela representam o PCA com os espectros sintéticos normalizados. Para comparação plotamos as PCs do caso sem normalização usando linha pontilhada.

5.2.3 NGC 4210 - CALIFA 518

A galáxia NGC 4210, também é uma espiral com uma barra bem notável, assim como a NGC 0776 (Figura 5.13). Pelas figuras podemos notar que existem problemas com algumas zonas e esses problemas passaram para a síntese, fazendo com que os resultados dessas zonas ficassem díspares ao todo. Esses problemas passaram para o cubo final mesmo passando pelos *pipelines* de redução, marcação de zonas defeituosas e extensiva revisão à olho.

O *scree test* (Figura 5.14) apesar de seguir o mesmo padrão das demais parece ter uma distribuição de variâncias mais parecida entre as PCs para casos diferentes, mas como falamos a seguir, essa galáxia é um caso à parte.

A tomografia PCA (Figuras 5.15 e 5.16) foi capaz de identificar esses problemas isolando-os nas duas primeiras PCs principalmente, pela extrema variância que eles impõe sobre os espectros. Na quinta PC vemos também um exemplo de “linha larga” que também foi causado por defeitos em zonas com espectros defeituosos em determinados comprimentos de onda. Apesar desses problemas, vemos que para os espectros sintéticos parte deles já desapareceram (veja PC2 em 5.15 e compare com a PC2 em 5.16).

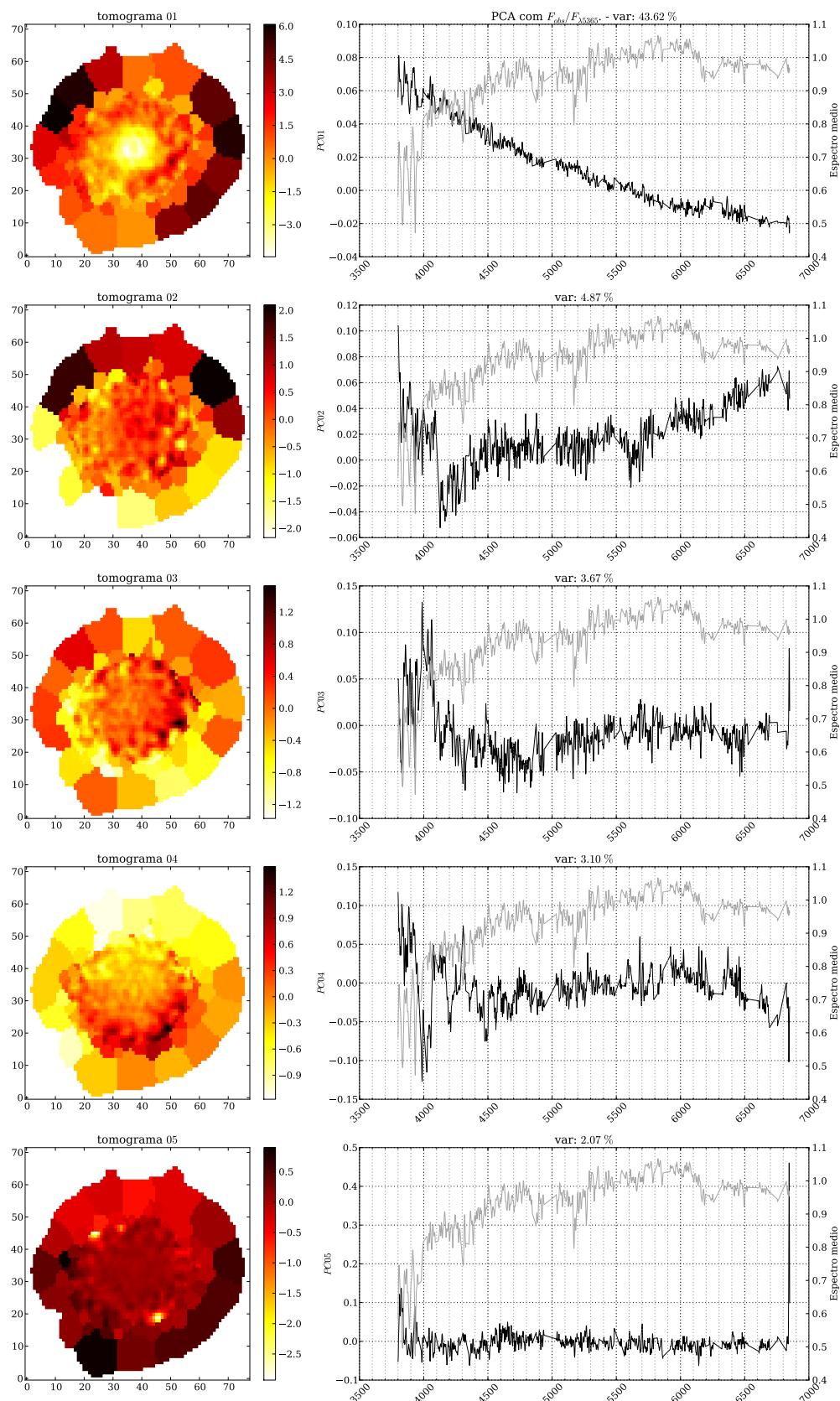


Figura 5.3: Cinco primeiras PCs (e seus respectivos tomogramas) resultantes da Tomografia PCA aplicado aos espectros observados com normalização da galáxia NGC 0001.

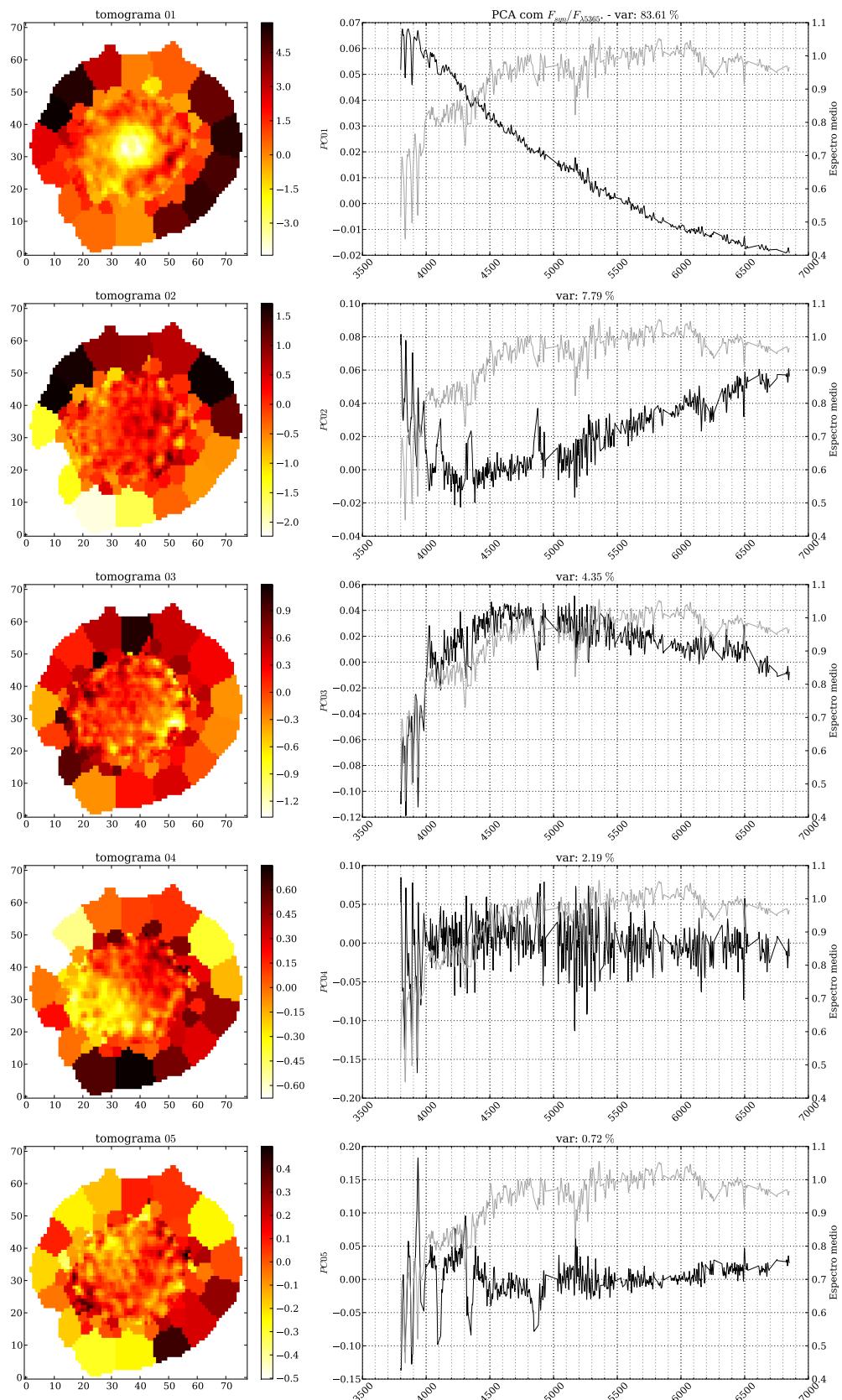


Figura 5.4: Cinco primeiras PCs (e seus respectivos tomogramas) resultantes da Tomografia PCA aplicado aos espectros sintéticos com normalização da galáxia NGC 0001.

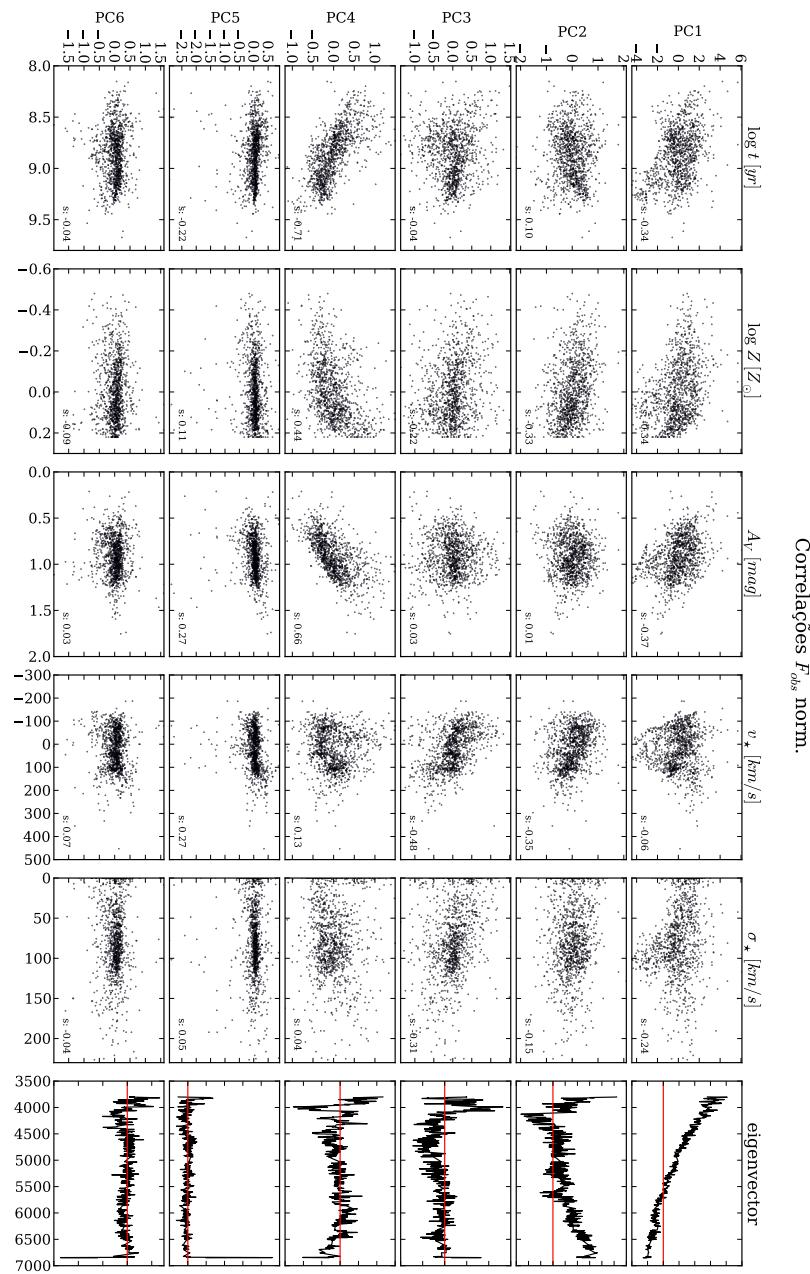


Figura 5.5: Correlações entre os pesos por zona das seis primeiras PCs do PCA feito para o cubo com os espectros observados normalizados e cinco parâmetros físicos. Pela ordem de colunas da esquerda para direita temos $\log t$, $\log Z$, A_V , v_* , σ_* . Na última coluna temos o autoespectro para ajudar na visualização. A linha em vermelho no gráfico do autoespectro serve para identificar o zero. O número dentro de cada gráfico é o coeficiente de correlação de Spearman.

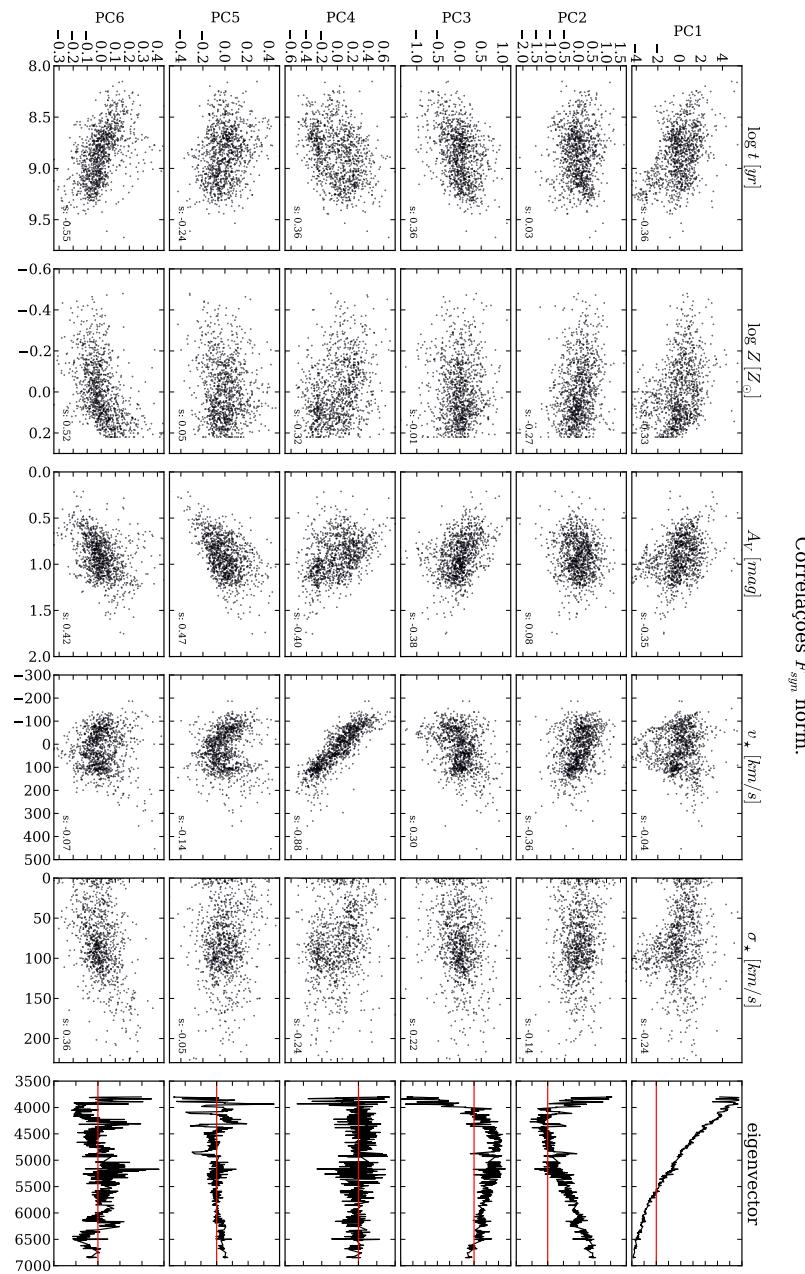
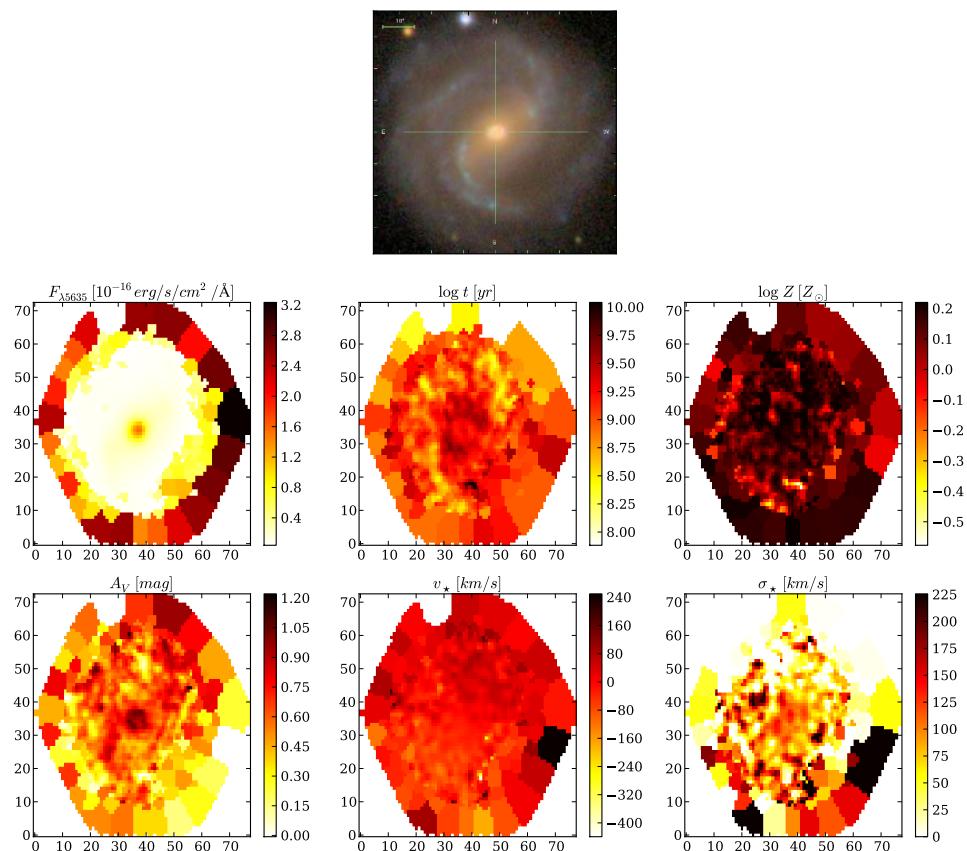


Figura 5.6: Correlações entre os pesos por zona das seis primeiras PCs do PCA feito para o cubo com os espectros sintéticos normalizados e cinco parâmetros físicos. Pela ordem de colunas da esquerda para direita temos $\log t$, $\log Z/Z_\odot$, A_V , v_\star , σ_\star . Na última coluna temos o autoespectro para ajudar na visualização. A linha em vermelho no gráfico do autoespectro serve para identificar o zero. O número dentro de cada gráfico é o coeficiente de correlação de Spearman.

NGC0776 - K0073

**Figura 5.7:** Igual a Figura 5.1 para a galáxia NGC 0776.

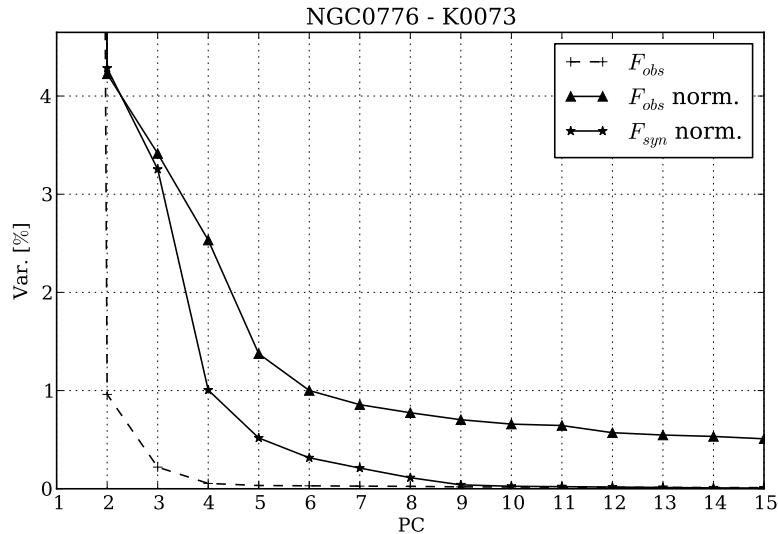


Figura 5.8: Igual a Figura 5.2 para a galáxia NGC 0776.

Com esses problemas, o surgimento dessa PC2 no caso observado faz com que fique um pouco mais complicado estudar o sentido físico dessas PCs, mas olhando as correlações para o PCA com os espectros observados normalizados (Figura 5.17) vemos que a PC4 é uma boa medida da velocidade ($s = 0.80$). Através das correlações usando o resultado para os espectros sintéticos (Figura 5.18) podemos ver que segue o padrão para outras galáxias espirais, onde a PC1 está geralmente fortemente correlacionada a idade. A PC4, da mesma forma que o observado, reproduz v_* e a PC5 mais próximo de A_V .

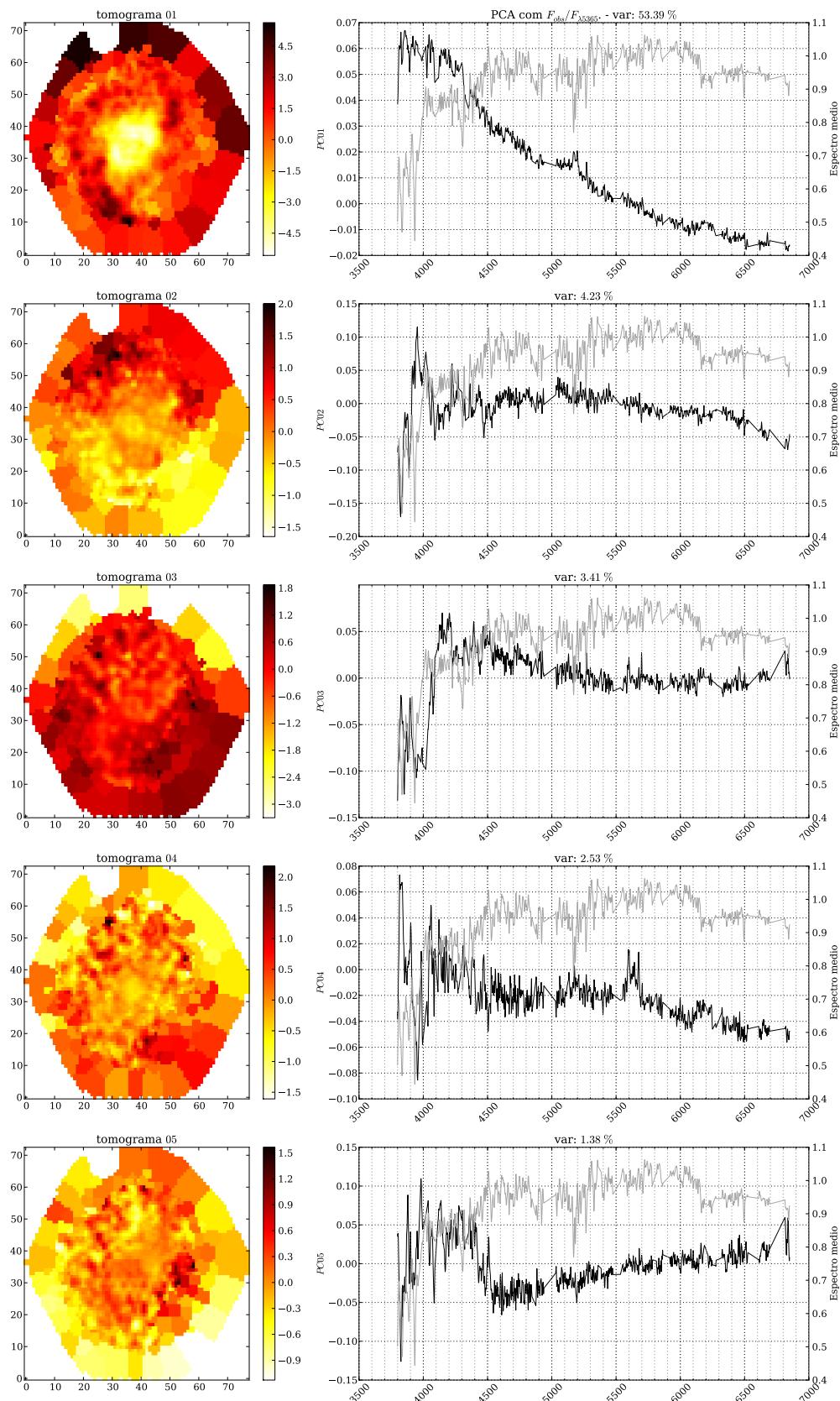


Figura 5.9: Igual a Figura 5.3 para a galáxia NGC 0776.

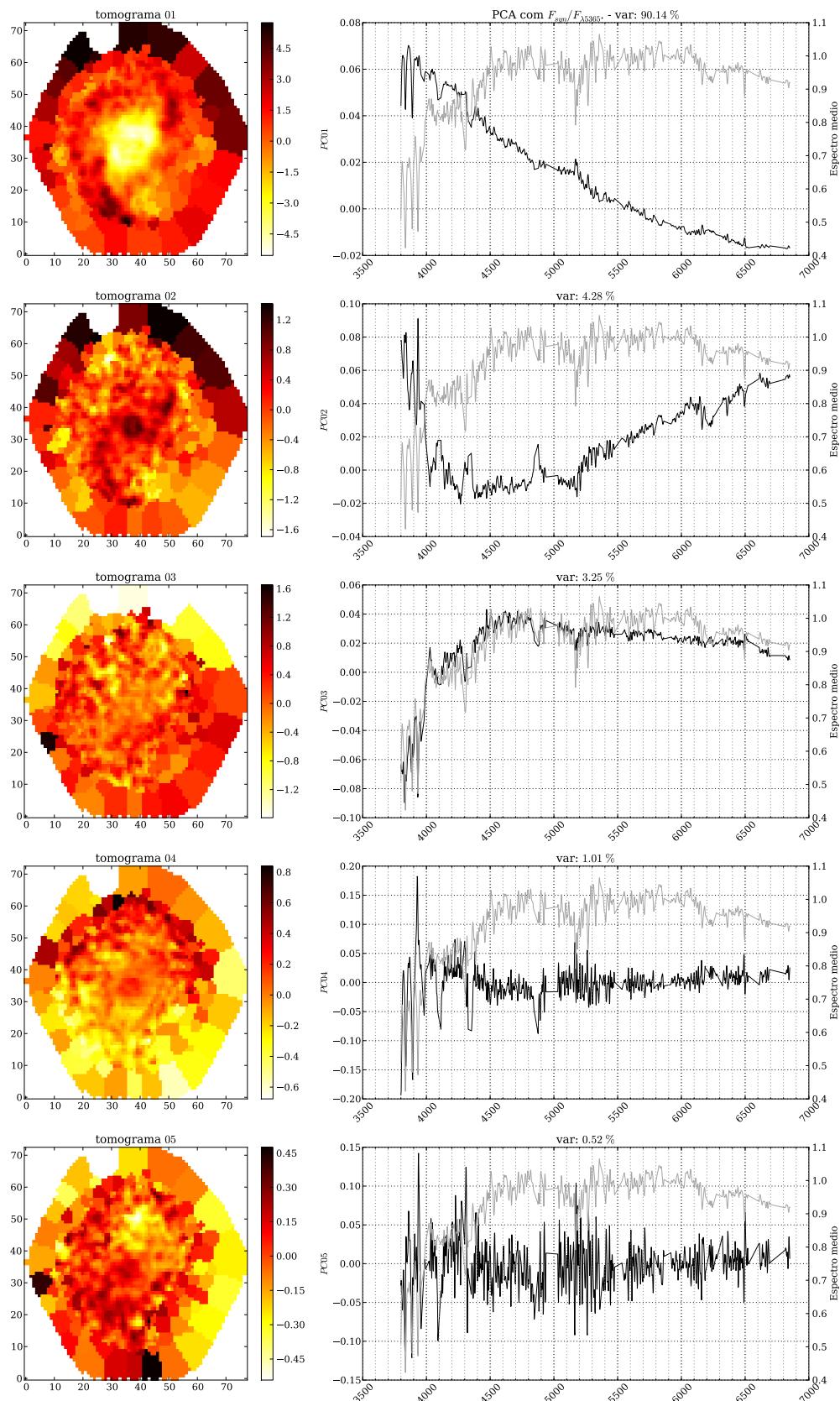


Figura 5.10: Igual a Figura 5.4 para a galáxia NGC 0776.

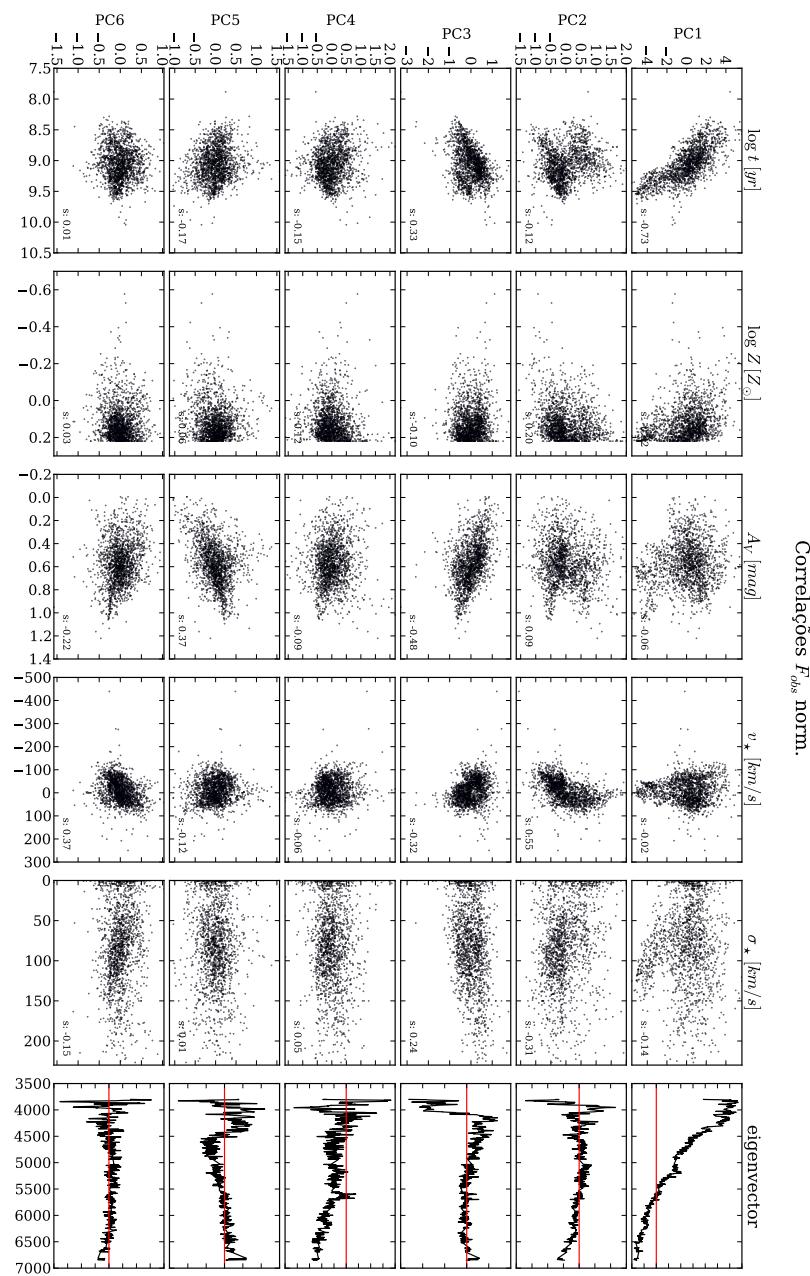


Figura 5.11: Igual a Figura 5.5 para a galáxia NGC 0776.

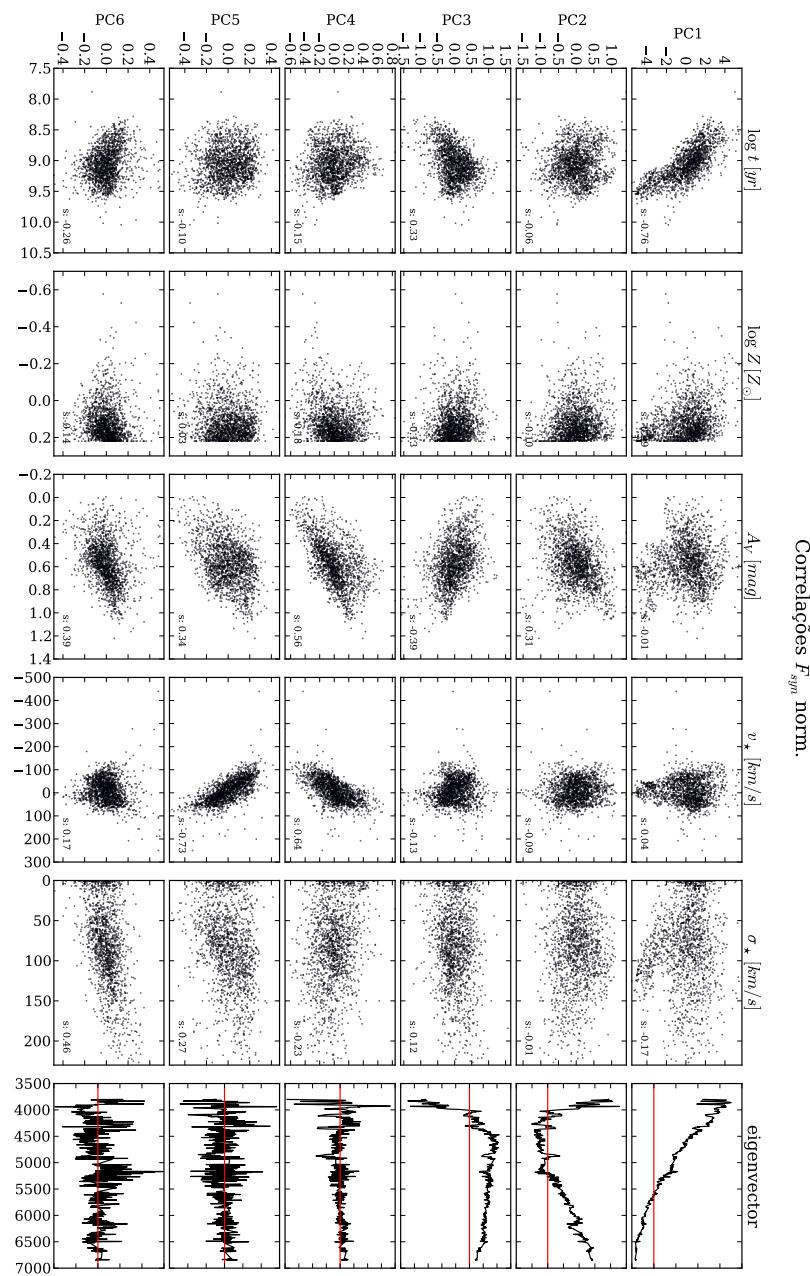


Figura 5.12: Igual a Figura 5.6 para a galáxia NGC 0776.

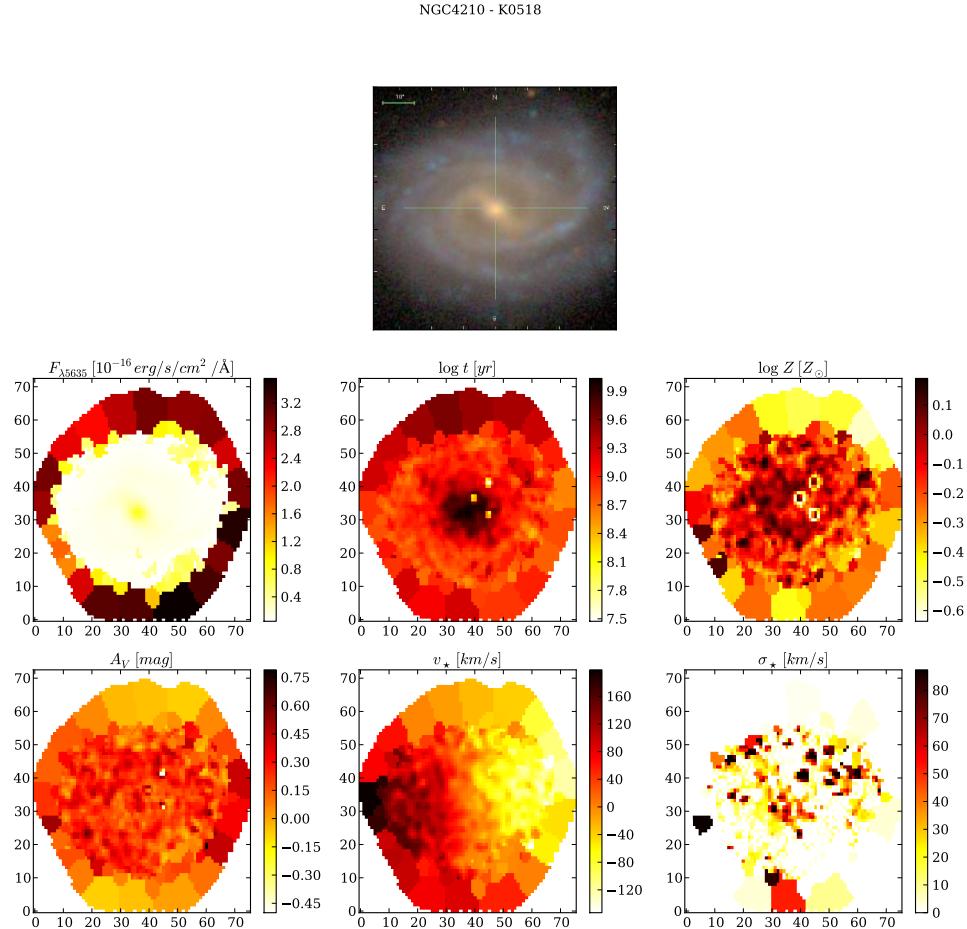


Figura 5.13: Igual a Figura 5.1 para a galáxia NGC 4210.

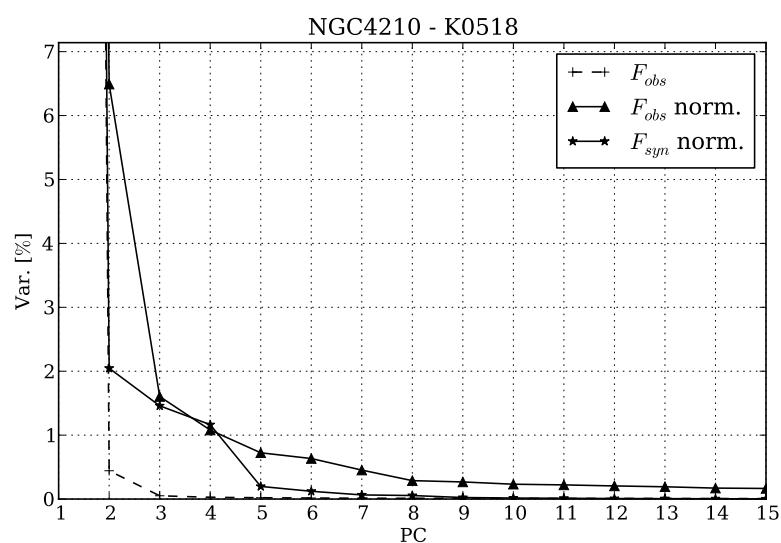


Figura 5.14: Igual a Figura 5.2 para a galáxia NGC 4210.

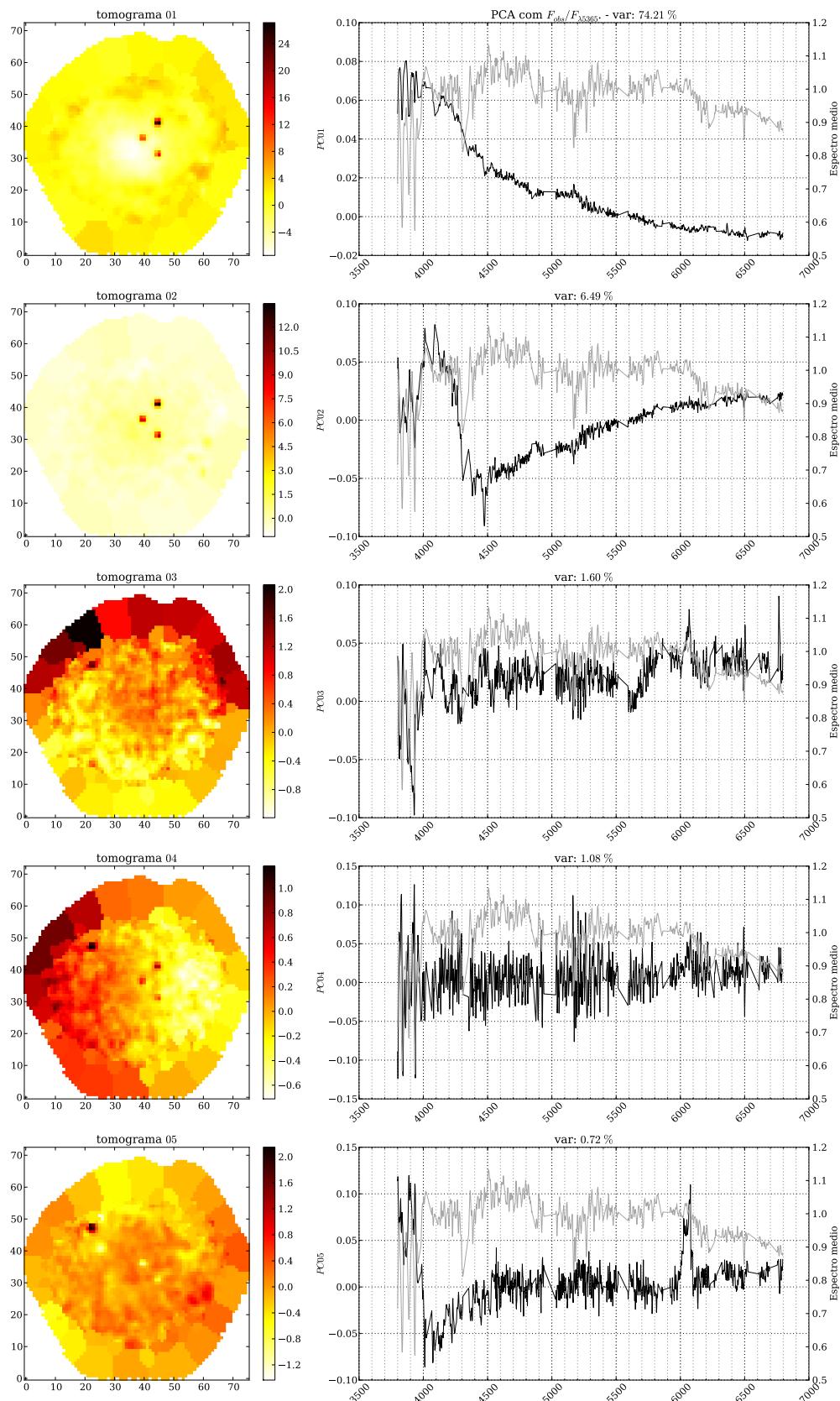


Figura 5.15: Igual a Figura 5.3 para a galáxia NGC 4210.

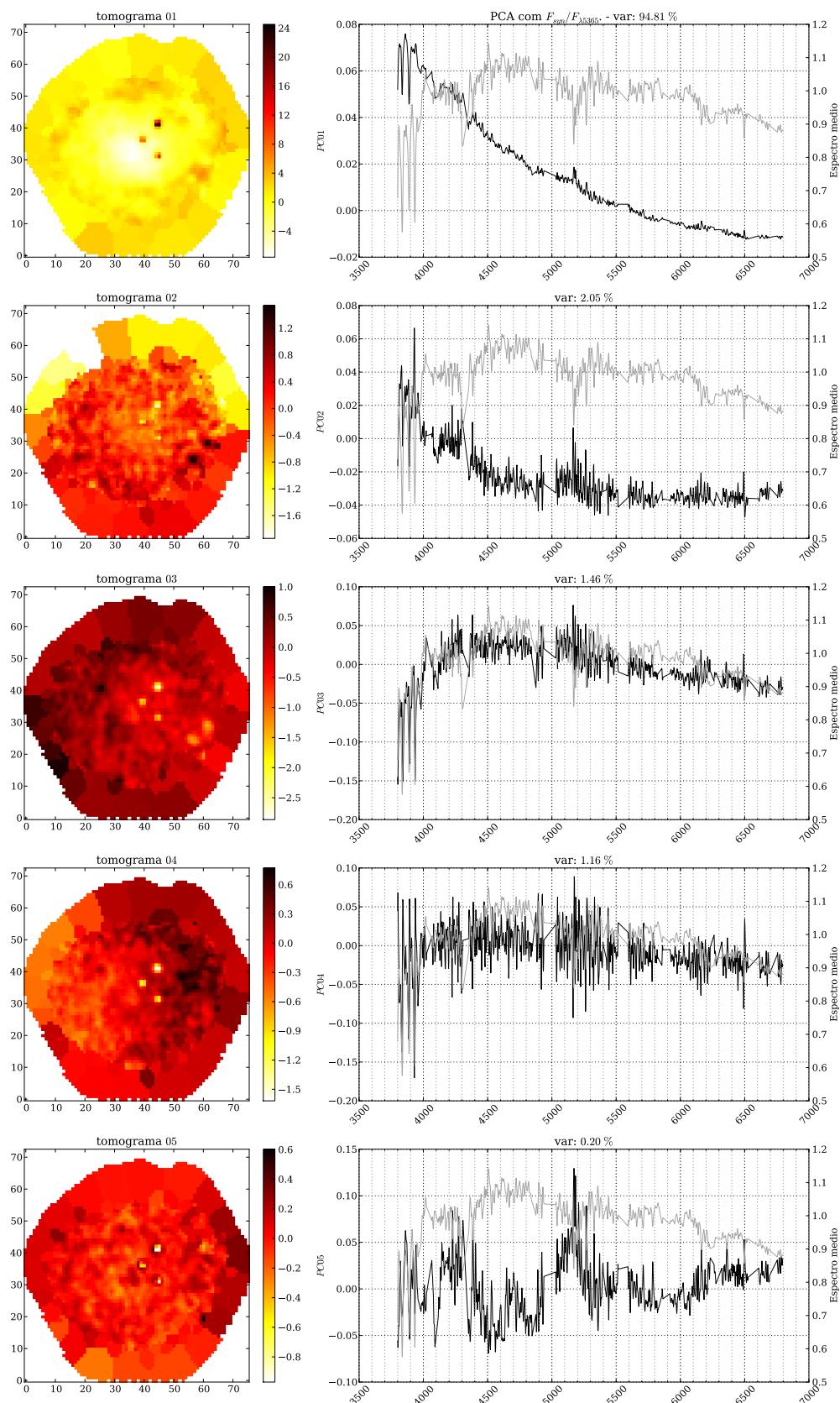


Figura 5.16: Igual a Figura 5.4 para a galáxia NGC 4210.

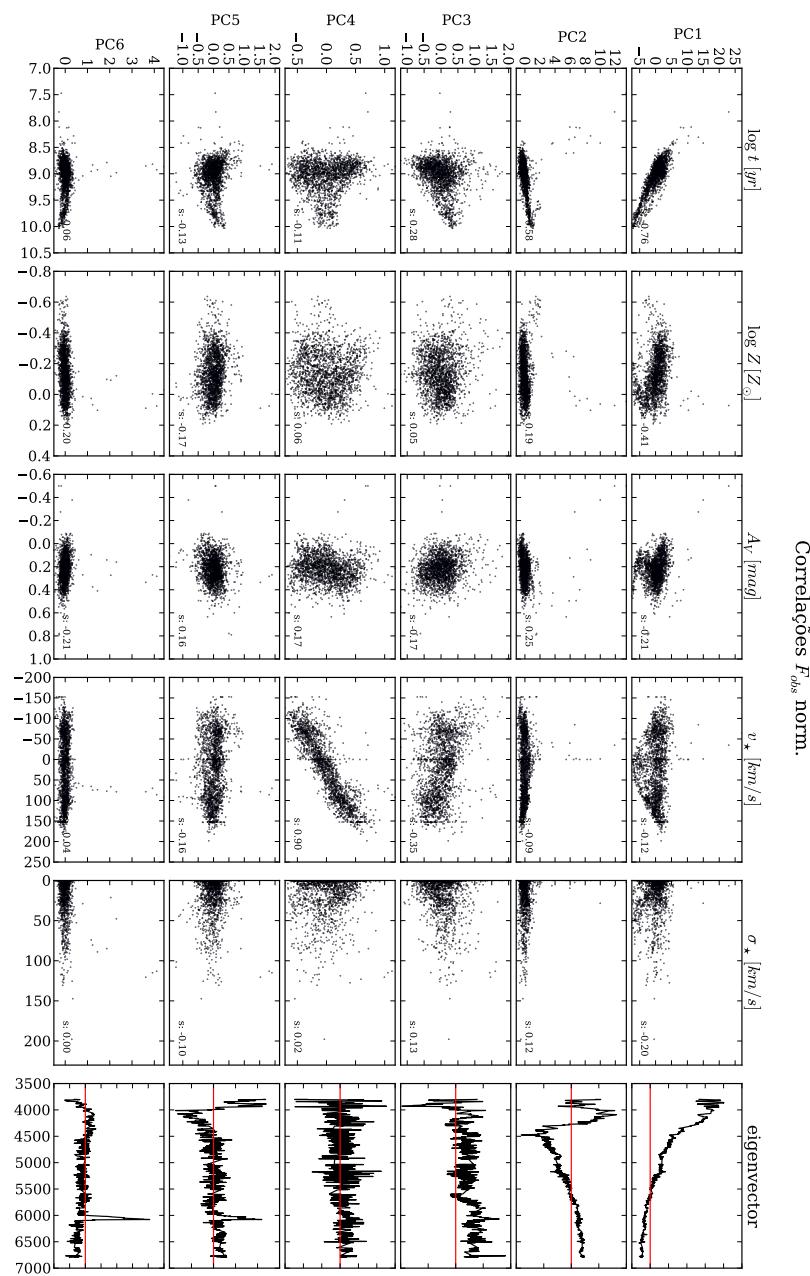


Figura 5.17: Igual a Figura 5.5 para a galáxia NGC 4210.

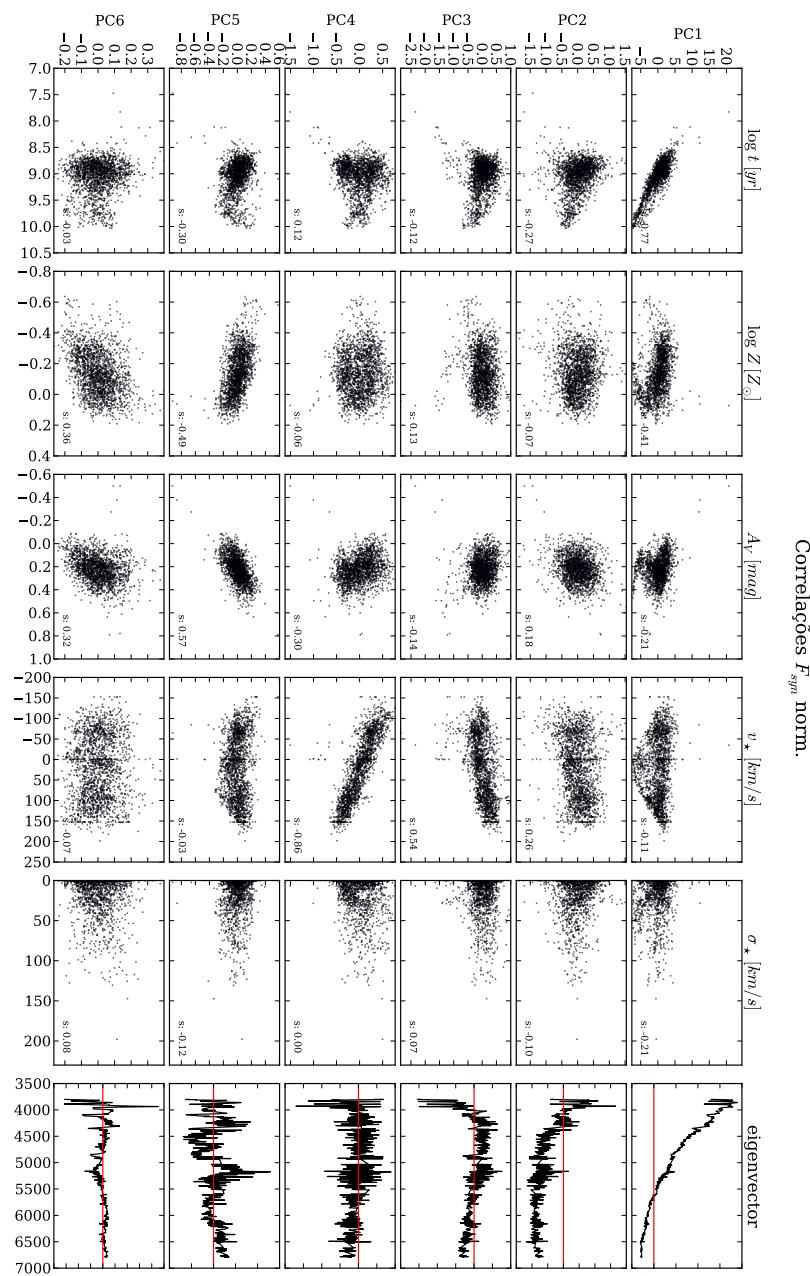


Figura 5.18: Igual a Figura 5.6 para a galáxia NGC 4210.

5.3 Galáxias *early-type*

??

5.3.1 NGC 1167 - CALIFA 119

Pela apresentação das propriedades físicas dessa galáxia (Figura 5.19), cores na imagem SDSS, distribuição de idades, e grande disperção de velocidades no centro, já mostram o quanto essa galáxia é diferente de todas que exploramos até agora.

No *scree test* o comportamento assintótico é o mesmo para as outras galáxias, com a variância mais distribuída entre as primeiras PCs no caso sintético. A segunda componente do caso sintético nem aparece no gráfico, mas pelo seu tomograma (Figura 5.22) vemos que ela possui $\sim 35\%$ da variância.

Vemos que a PC1 dos dois casos (Figuras 5.21 e 5.22) são bem diferentes e no caso sintético ela não é correlacionada com alguma propriedade específica. Já no caso observado tem uma ótima correlação com A_V e com a metalicidade (!@j@! até agora não falei de metalicidade em lugar algum, mas não lembro de ver uma correlação tão grande com metalicidade). No sintético A_V aparece principalmente na PC2, também misturada com a metalicidade. A PC5 do caso observado também correlaciona bem com A_V . O padrão de velocidades aparece de forma muito clara para ambos os casos na PC3, com um inclível índice de Spearmann de 0.98 para o caso sintético. Vemos que a disperção de velocidades (σ_\star) aparece correlacionando fracamente com quase todas as principais PCs, mas no caso sintético vemos uma correlação maior ($s = 0.53$). (!@j@! olhar a metalicidade e idade na PC5 do caso sintético, correlação com metalicidade = 0.56.) A idade média estelar fica mais distribuída entre as PC5 e PC6 no caso sintético, mas mesmo assim com correlações não tão gritantes quanto para as galáxias espirais. Já no caso observado correlaciona um pouco com a PC2 mas nada muito evidente também.

5.3.2 NGC 6515 - CALIFA 864

A primeira coisa que chama atenção nas imagens produzidas pelas propriedades físicas dessa galáxia é o número de zonas faltantes na análise. Existem muitos objetos no *FoV* de observação dessa galáxia, portanto esses espaços são causados pelas máscaras espaciais criadas pelo

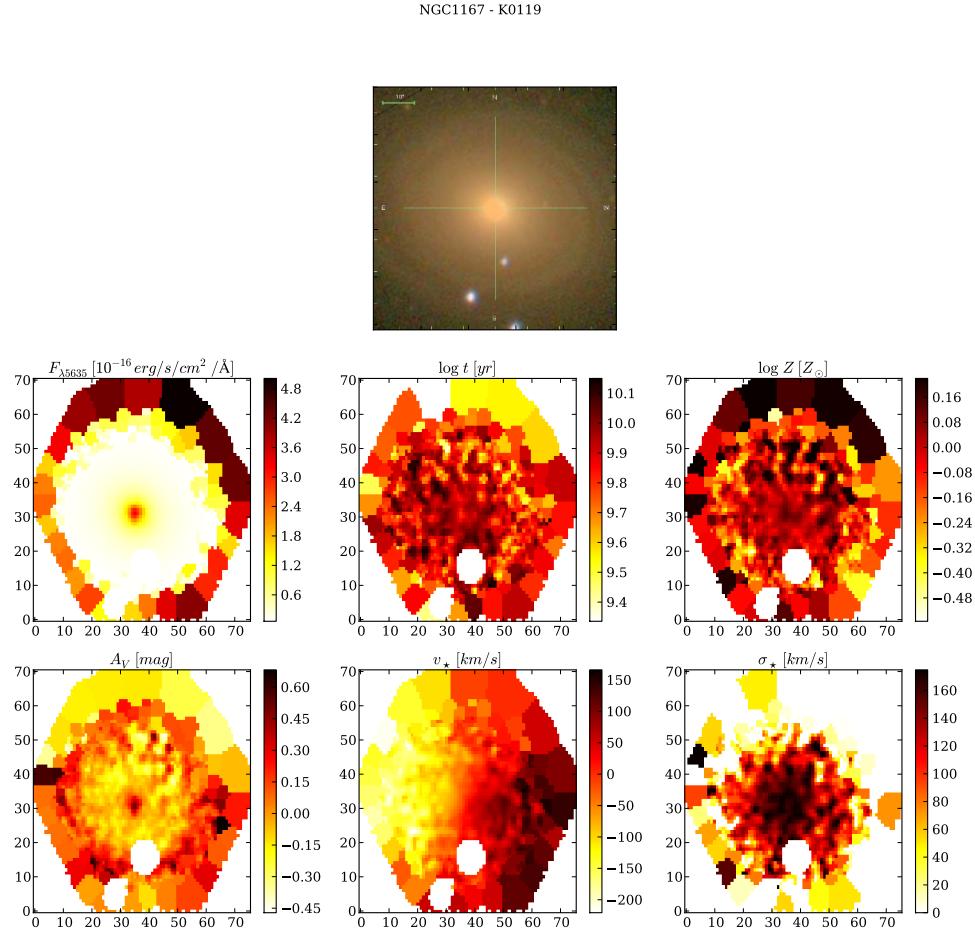


Figura 5.19: Igual a Figura 5.1 para a galáxia NGC 1167.

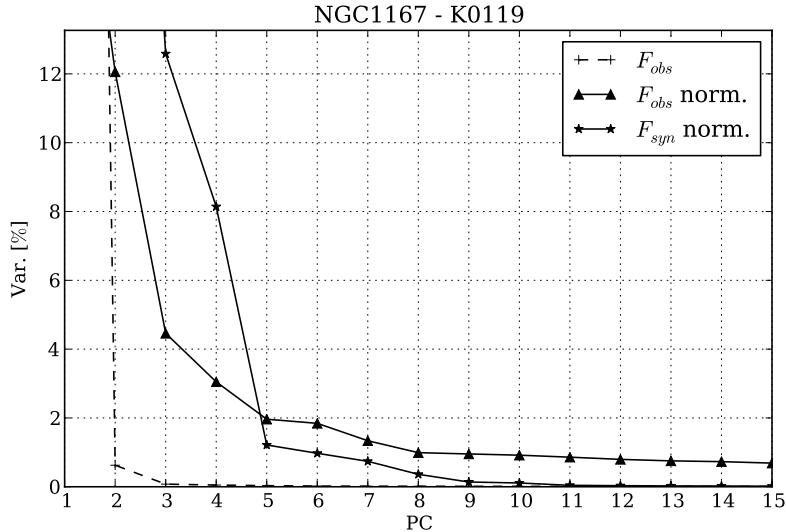


Figura 5.20: Igual a Figura 5.2 para a galáxia NGC 1167.

qbick conforme apresentado na Seção ???. O *scree test* (Figura 5.26) possui o mesmo resultado assintótico das outras galáxias, mas mostra uma variância percentual alta mesmo para componentes mais afastadas das cinco principais.

Analizando os tomogramas e PCs para o caso observado (Figura 5.27) não parecem refletir comportamentos físicos bem definidos. Podemos ver através das correlações (Figura 5.29) que não está claro que exista alguma correlação singular, salvo PC1 e talvez a PC3. A PC1 correlaciona com A_V e a PC3 com a dispersão de velocidades (σ_*), mas nada muito notável. Para o caso sintético (tomogramas e PCs na Figura 5.27 e correlações na Figura 5.30) temos uma correlação da PC1 com A_V também, mas temos umas correlações mais fortes. A PC3 correlaciona com quase tudo, o que mostra que ela deve ser um tipo de fator de escala que segue o mesmo gradiente das demais propriedades físicas. Nessa PC ainda podemos ver que para a disperção de velocidades a correlação é um pouco mais alta ($s = 0.62$). A PC4 possui o mesmo padrão de correlações distribuidas entre as propriedades, mas é mais notável a correlação com a velocidade estelar (v_*), o que fica claro também observando a PC (autoespectro) que possui um padrão de oscilação em torno do zero. Detalhe para a PC6 que possui forte correlação com idade e metalicidade.

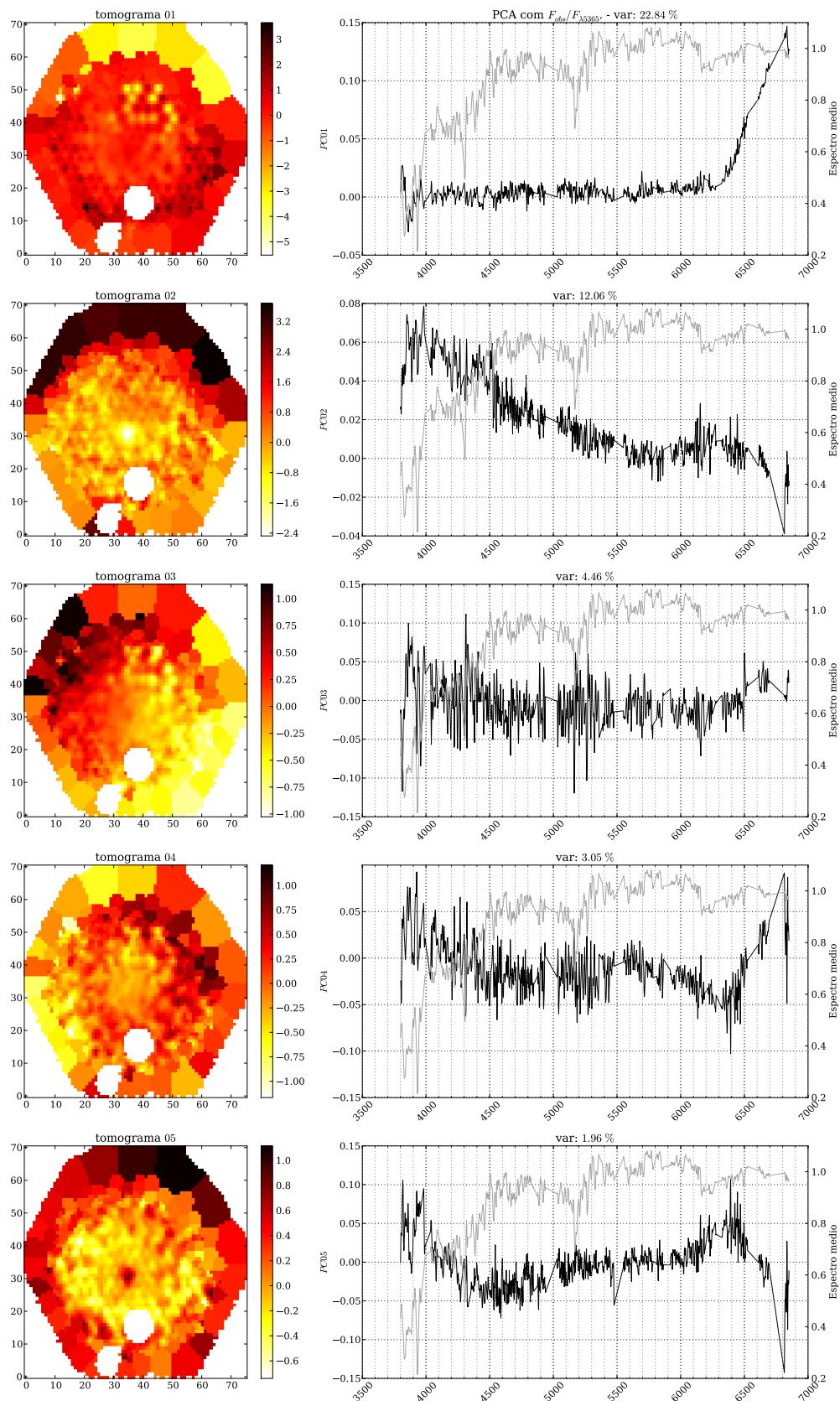


Figura 5.21: Igual a Figura 5.3 para a galáxia NGC 1167.

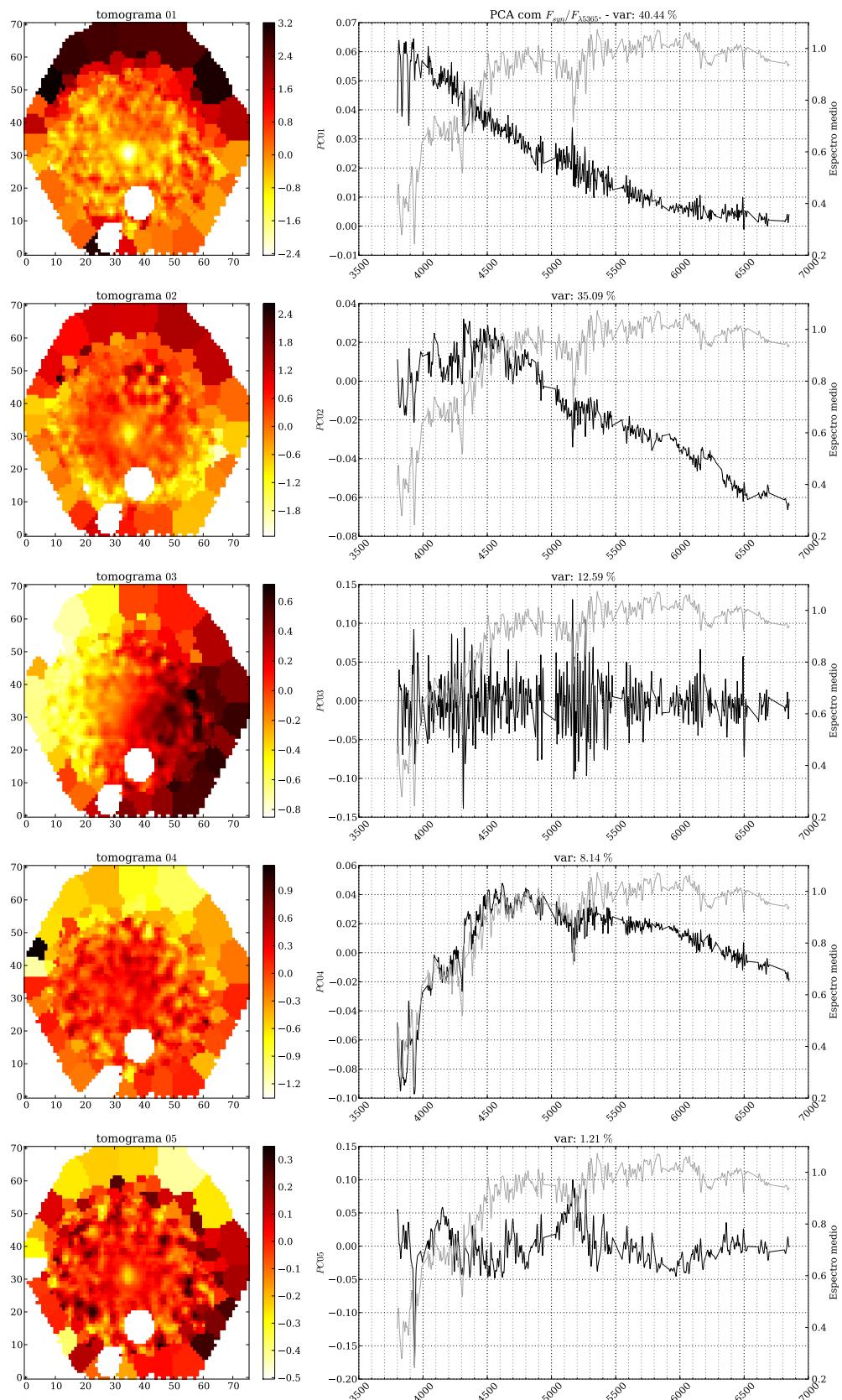


Figura 5.22: Igual a Figura 5.4 para a galáxia NGC 1167.

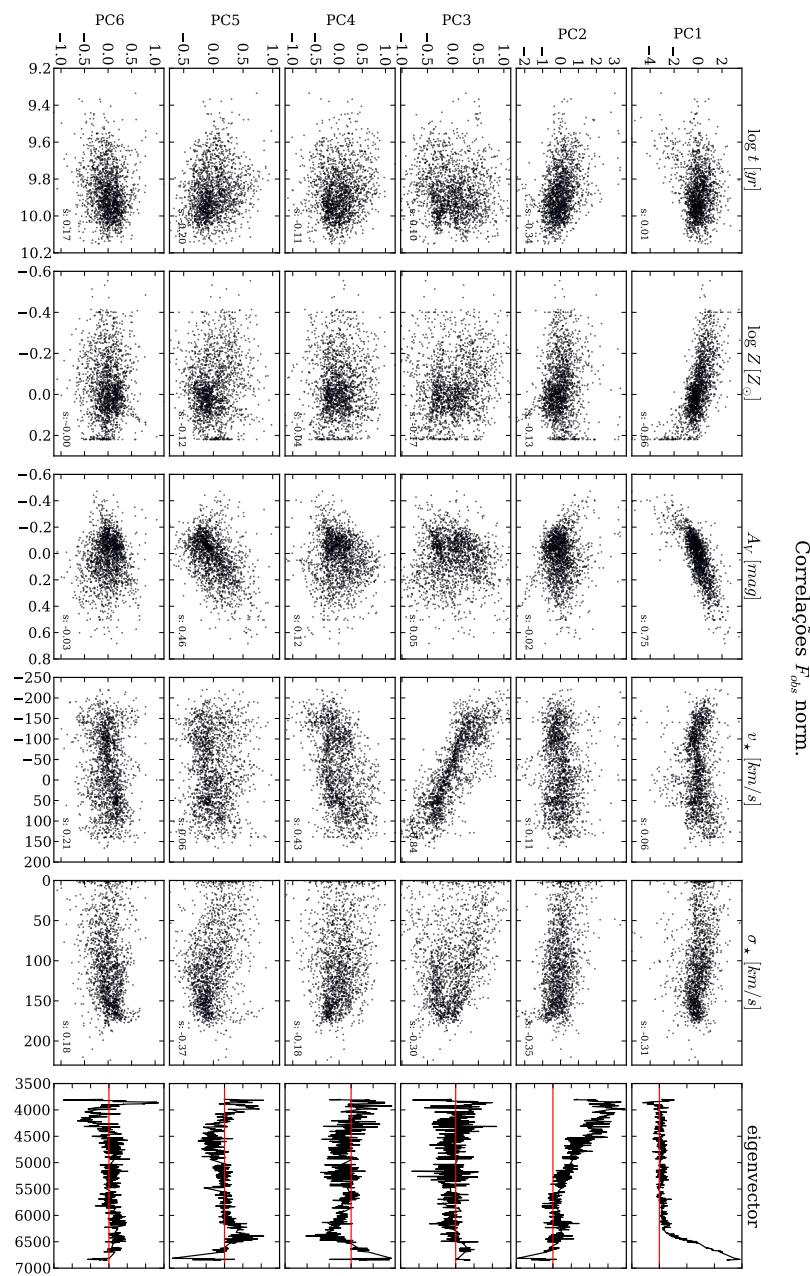


Figura 5.23: Igual a Figura 5.5 para a galáxia NGC 1167.

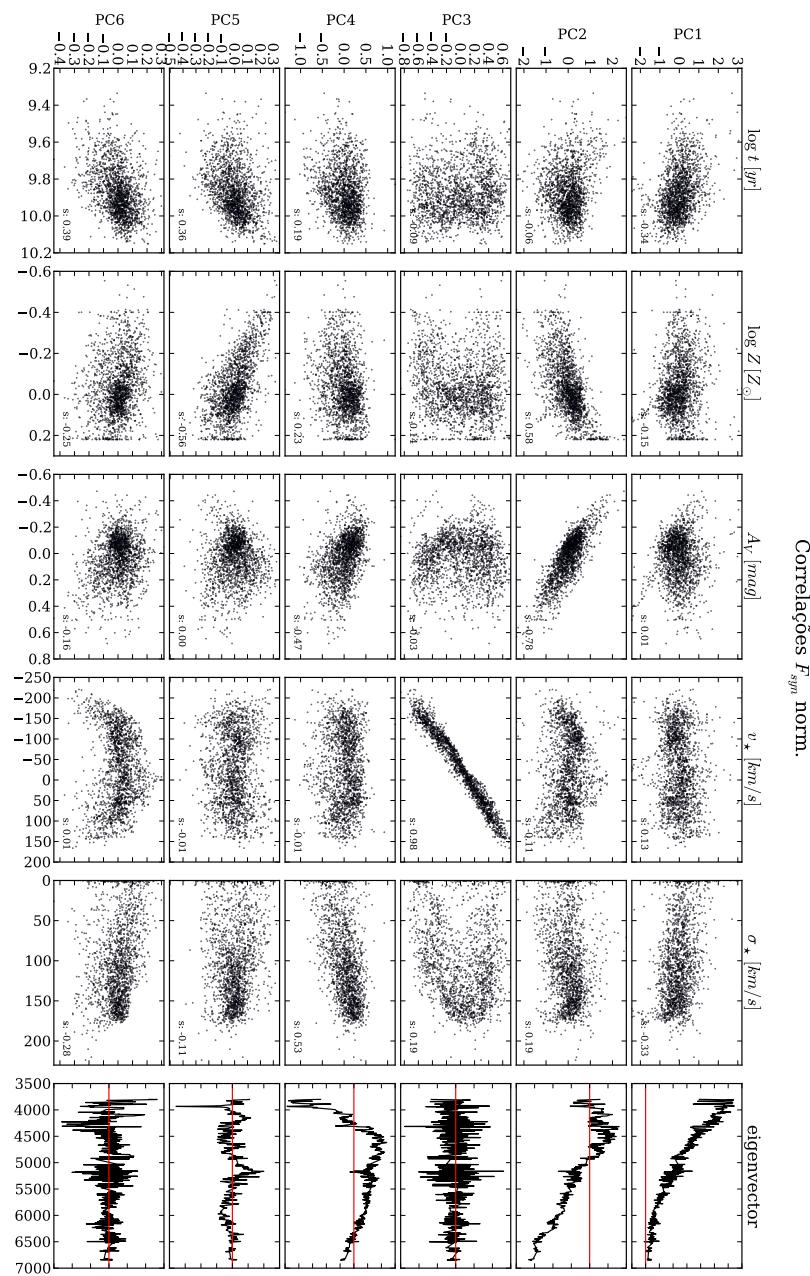


Figura 5.24: Igual a Figura 5.6 para a galáxia NGC 1167.

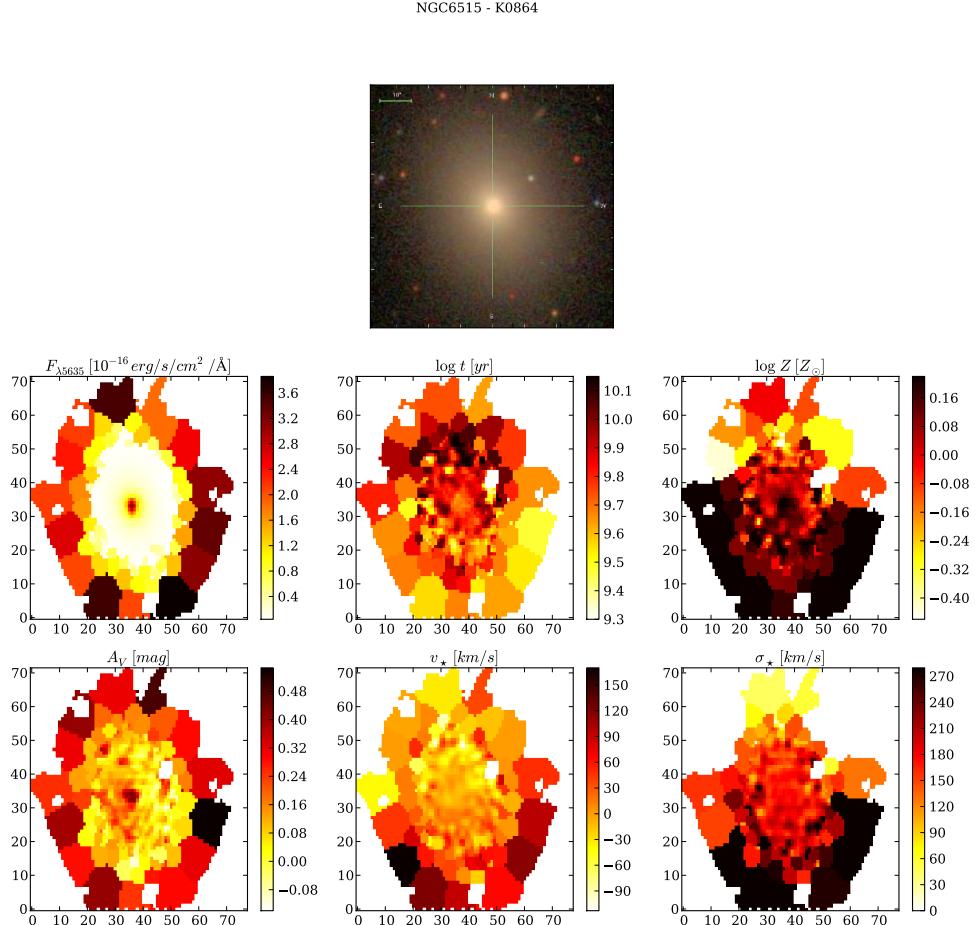


Figura 5.25: Igual a Figura 5.1 para a galáxia NGC 6515.

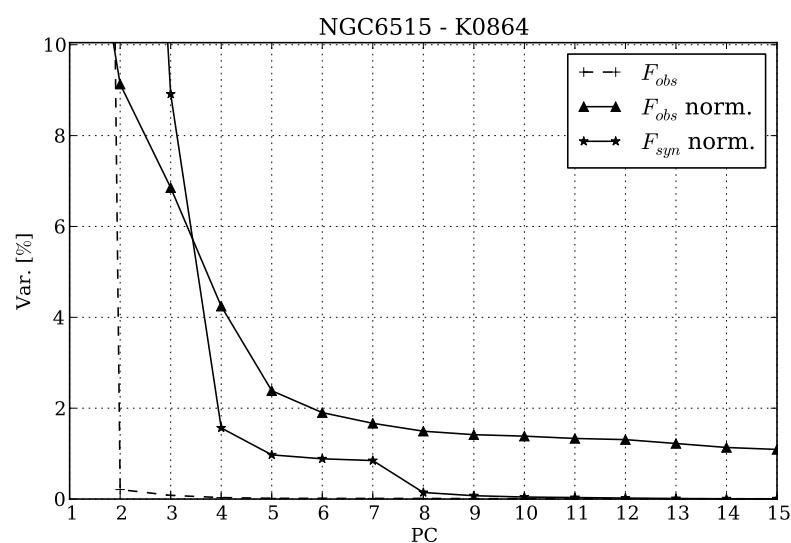


Figura 5.26: Igual a Figura 5.2 para a galáxia NGC 6515.

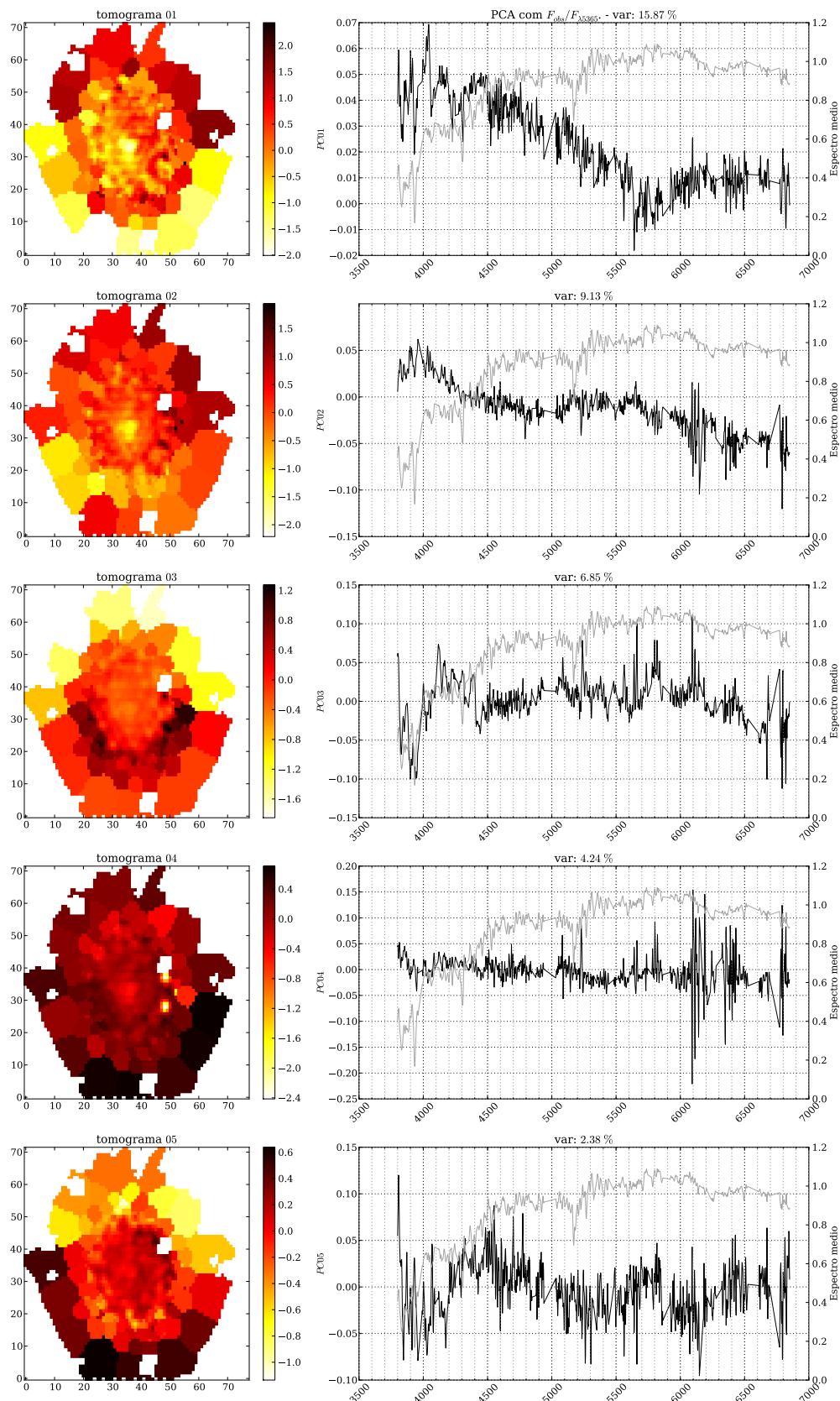


Figura 5.27: Igual a Figura 5.3 para a galáxia NGC 6515.

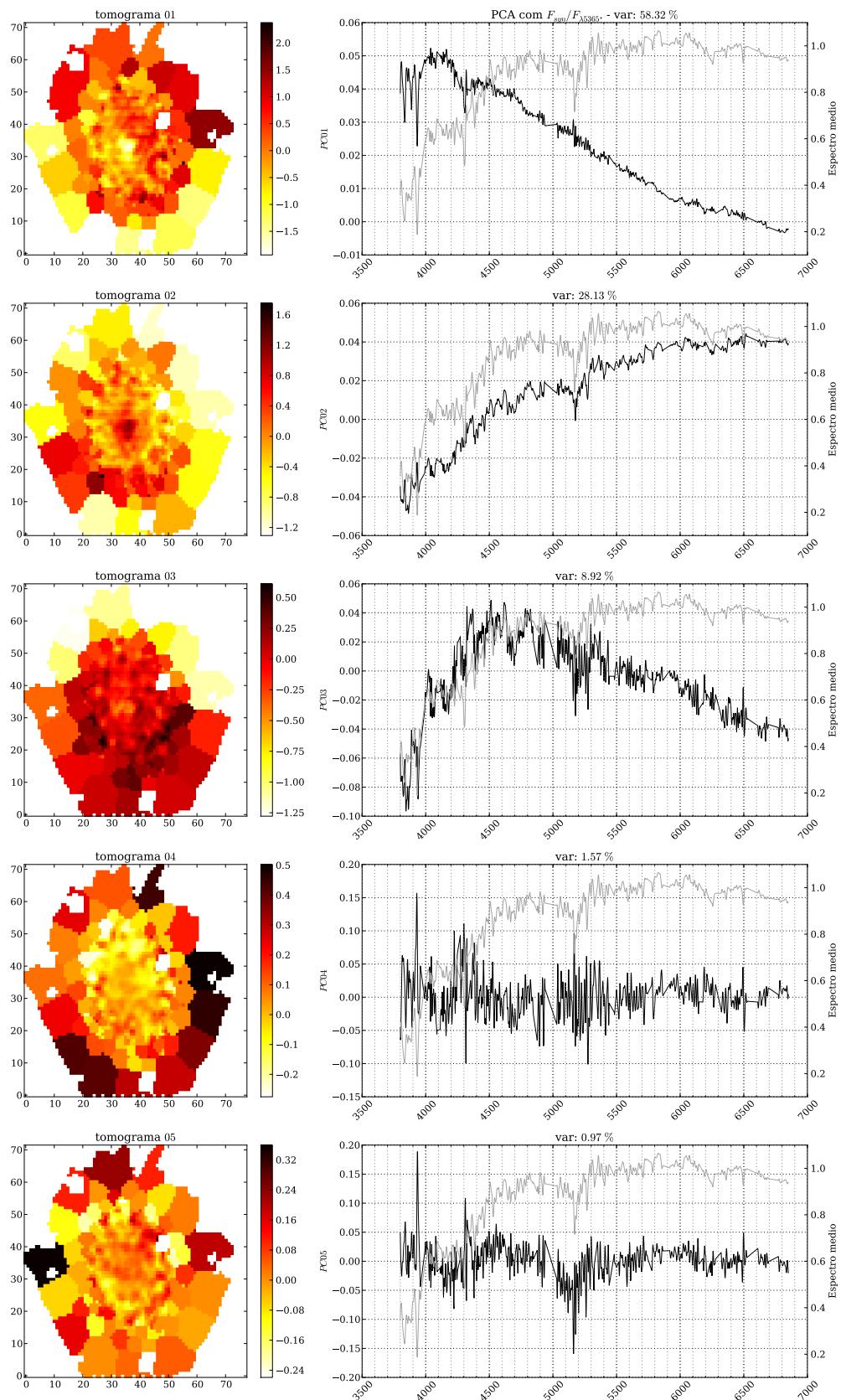


Figura 5.28: Igual a Figura 5.4 para a galáxia NGC 6515.

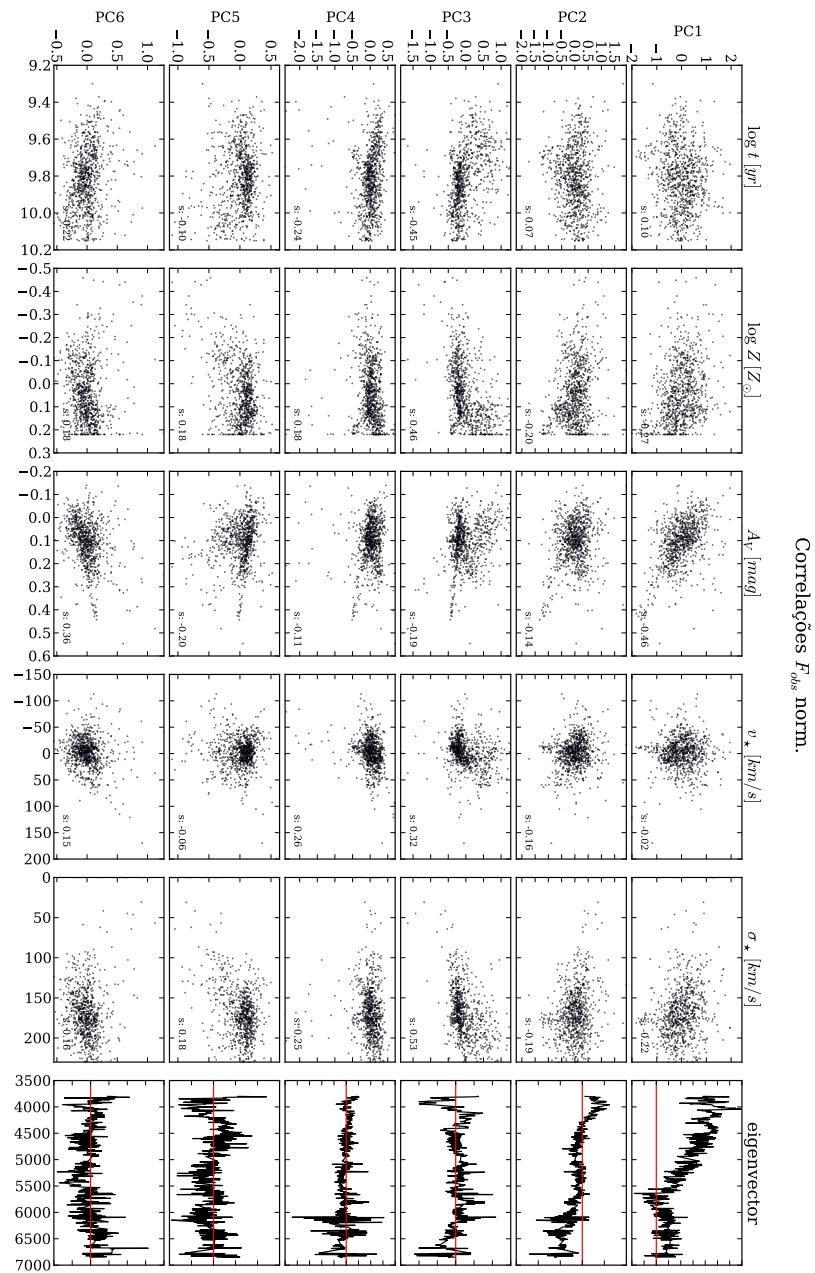


Figura 5.29: Igual a Figura 5.5 para a galáxia NGC 6515.

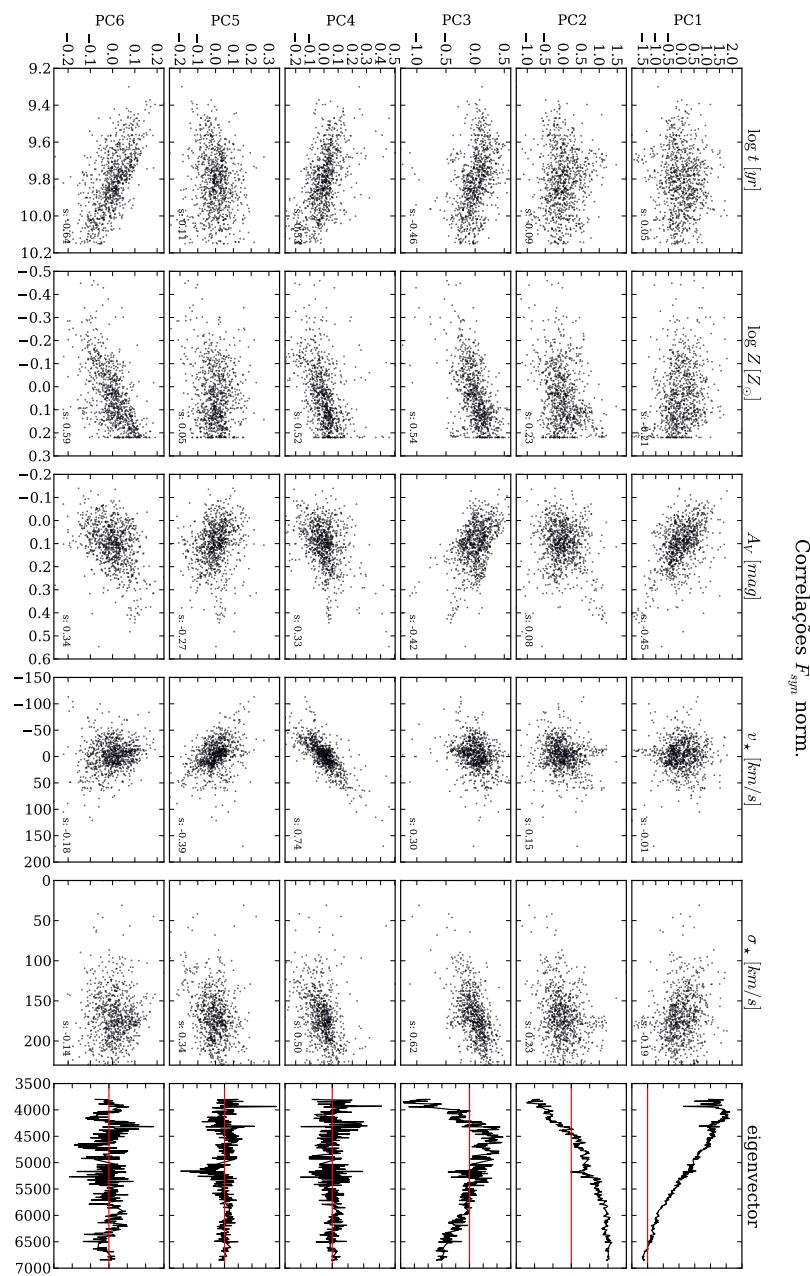


Figura 5.30: Igual a Figura 5.6 para a galáxia NGC 6515.

5.4 *Mergers*

??

5.4.1 NGC 2623 - CALIFA 213

Quando olhamos a figura de apresentação da imagem e das propriedades da galáxia NGC 2623 (Figura 5.31) fica evidente a natureza irregular e a diferença na distribuição radial da idade média das populações estelares. Seu *scree test* (Figura 5.32) possui o mesmo comportamento com as variâncias indo a zero mais rápido no caso sintético, mas surpreende pela alta concentração de informação em variância nas 4 primeiras componentes para o caso sintético.

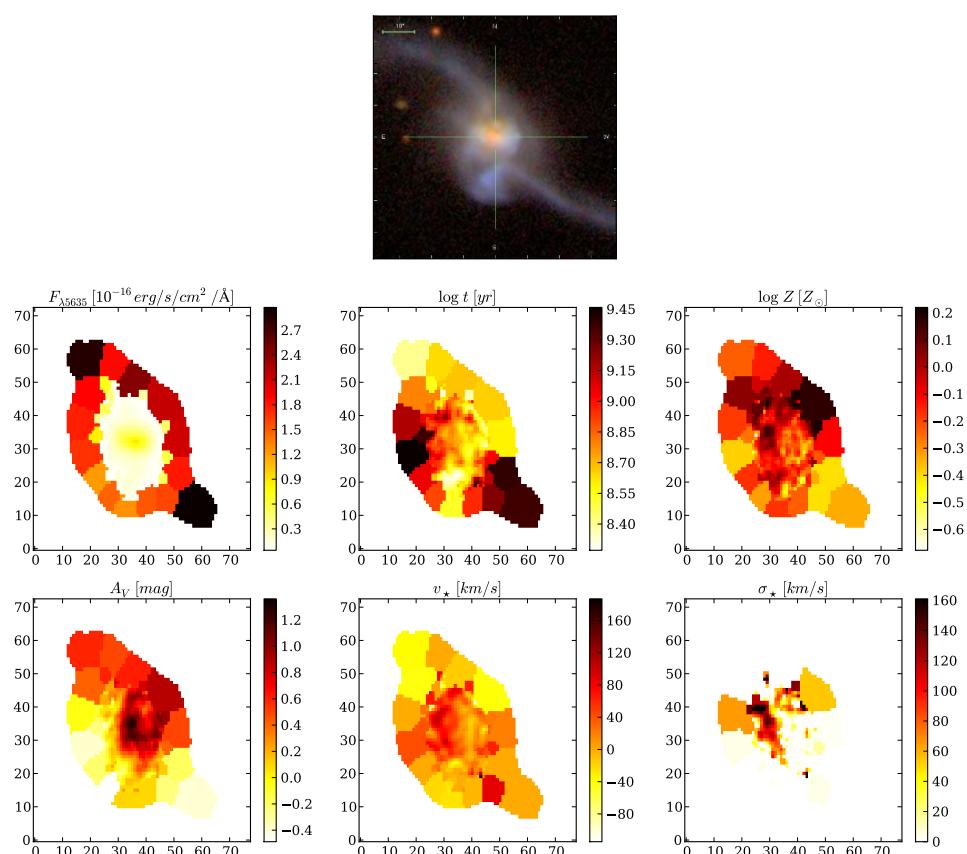
As duas primeiras PCs e tomogramas são muito semelhantes nos dois casos (Figuras 5.33 e 5.34) mas começam a diferir a partir do terceiro autovetor. A PC1 para os dois casos correlaciona muito bem com A_V e a segunda componente com correlação com todas as propriedades, mas de uma maneira mais forte com a idade (Figuras 5.35 e 5.36). A PC3 para o caso sintético não parece ter correlação com nenhuma das propriedades avaliadas. Já no caso observado vemos uma pequena correlação com a metalicidade. Chama atenção a sequência de balmer aparecendo na PC4 do caso sintético, mas tanto no caso sintético quanto no observado não parece correlacionar com nada. A quinta componente no caso observado parece ter algum vestígio de idade e metalicidade em ambos os casos, mas mais fortemente para o caso sintético. Com uma correlação um pouco maior, a velocidade estelar aparece na PC6 do caso sintético, mas não na do caso observado.

5.4.2 ARP 220 - CALIFA 802

Esse objeto peculiar é a ULIRG (*Ultraluminous infrared galaxy*) mais próxima da Terra. Como a NGC 2623, possui uma densa nuvem de poeira no centro (Figura 5.37). Repete o mesmo comportamento assintótico de todos os *scree tests* (Figura 5.38) anteriores e possui ~ 99% da variância contida nas primeiras 5 PCs para o caso observado (~ 99.7% para o caso sintético).

Também como a NGC 2623, possui as duas primeiras PCs muito semelhantes (Figuras 5.39 e 5.40). Comparando as correlações (Figuras 5.41 e 5.42) vemos que a primeira PC correlaciona fortemente com A_V e de maneira mais fraca com a idade. A segunda, apesar de ter uma leve correlação com todos as propriedades, a metalicidade se ressalta em ambos os casos.

NGC2623 - K0213

**Figura 5.31:** Igual a Figura 5.1 para a galáxia NGC 2623.

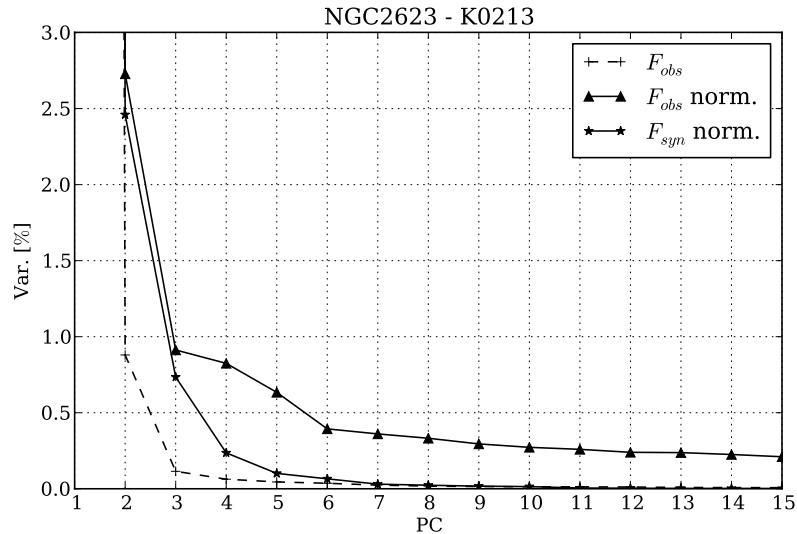


Figura 5.32: Igual a Figura ?? para a galáxia NGC 2623.

Para o caso observado a PC3 correlaciona melhor com a velocidade estelar. No caso sintético vemos vestígios da sequencia de Balmer tanto na PC3 (que não correlaciona fortemente com nada) quanto na PC4 (que parece ser uma boa medida de v_*). Ainda no caso sintético, vemos que a velocidade aparece fortemente também na PC5, mas divide espaço (em correlação) com idade e metalicidade. de maneira mais fraca. Já para o caso observado nenhuma delas (PC4, PC5) parecem ter correlação com nenhum dos parâmetros físicos comparados. O mesmo para a PC6 em ambos os casos.

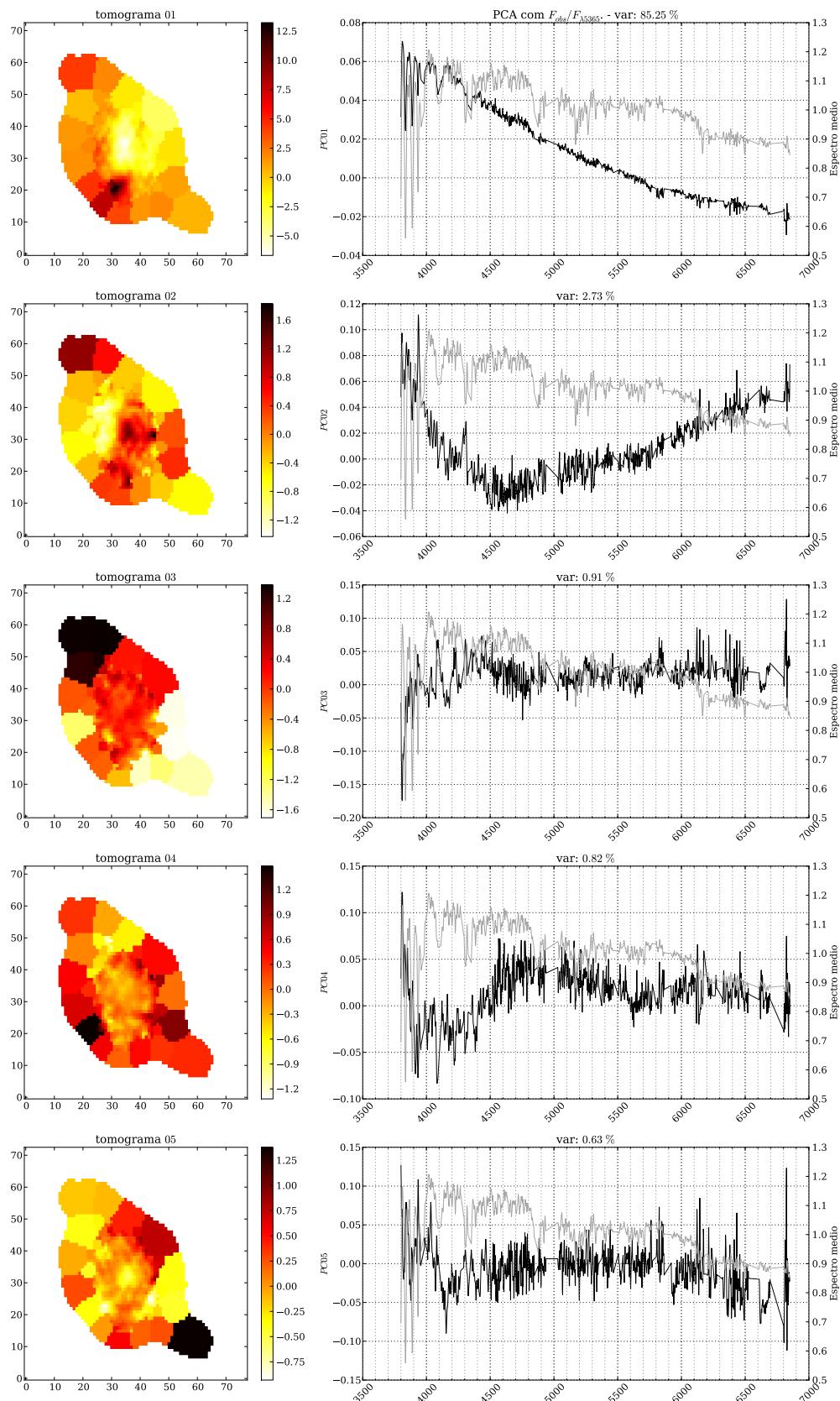


Figura 5.33: Igual a Figura 5.3 para a galáxia NGC 2623.

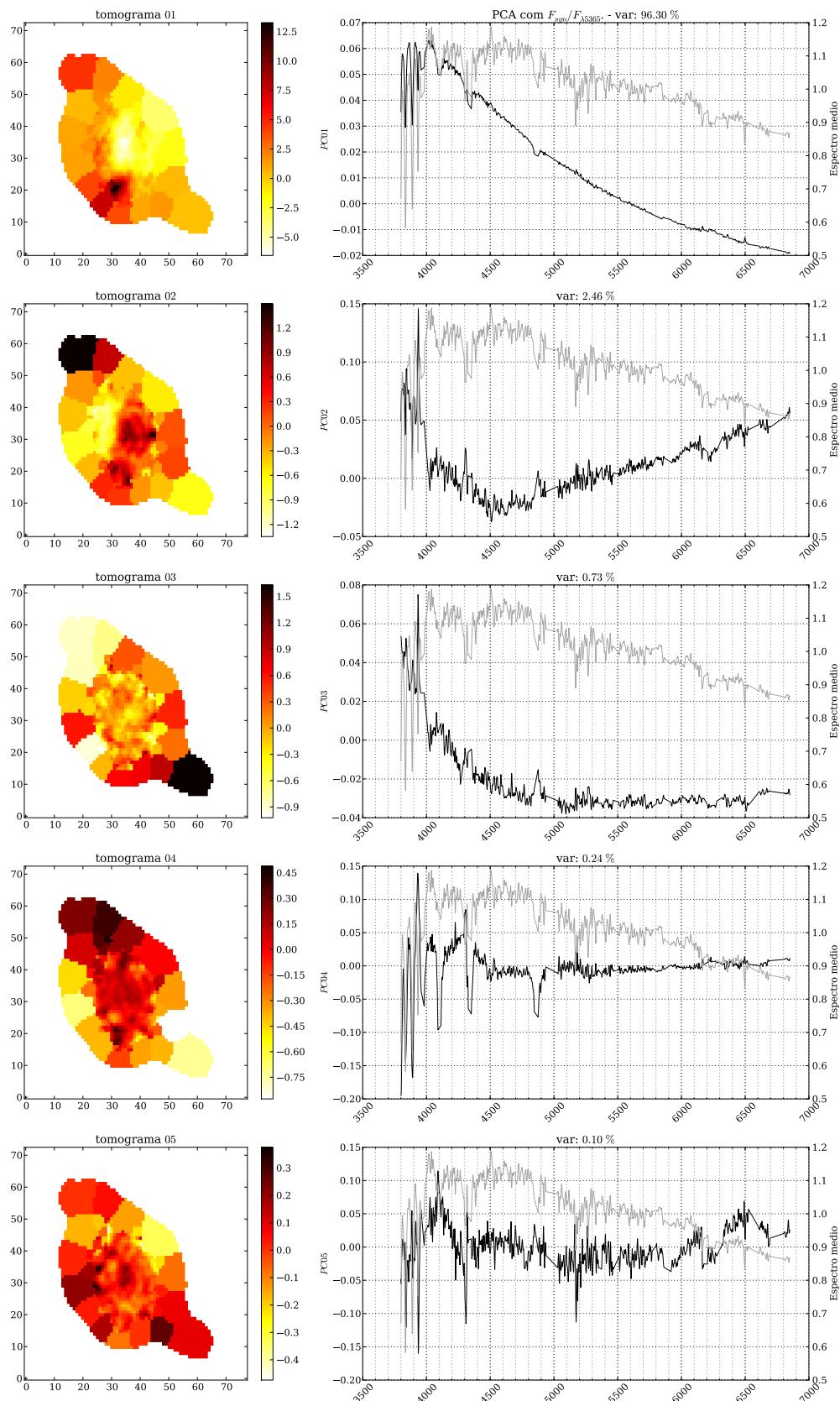


Figura 5.34: Igual a Figura 5.4 para a galáxia NGC 2623.

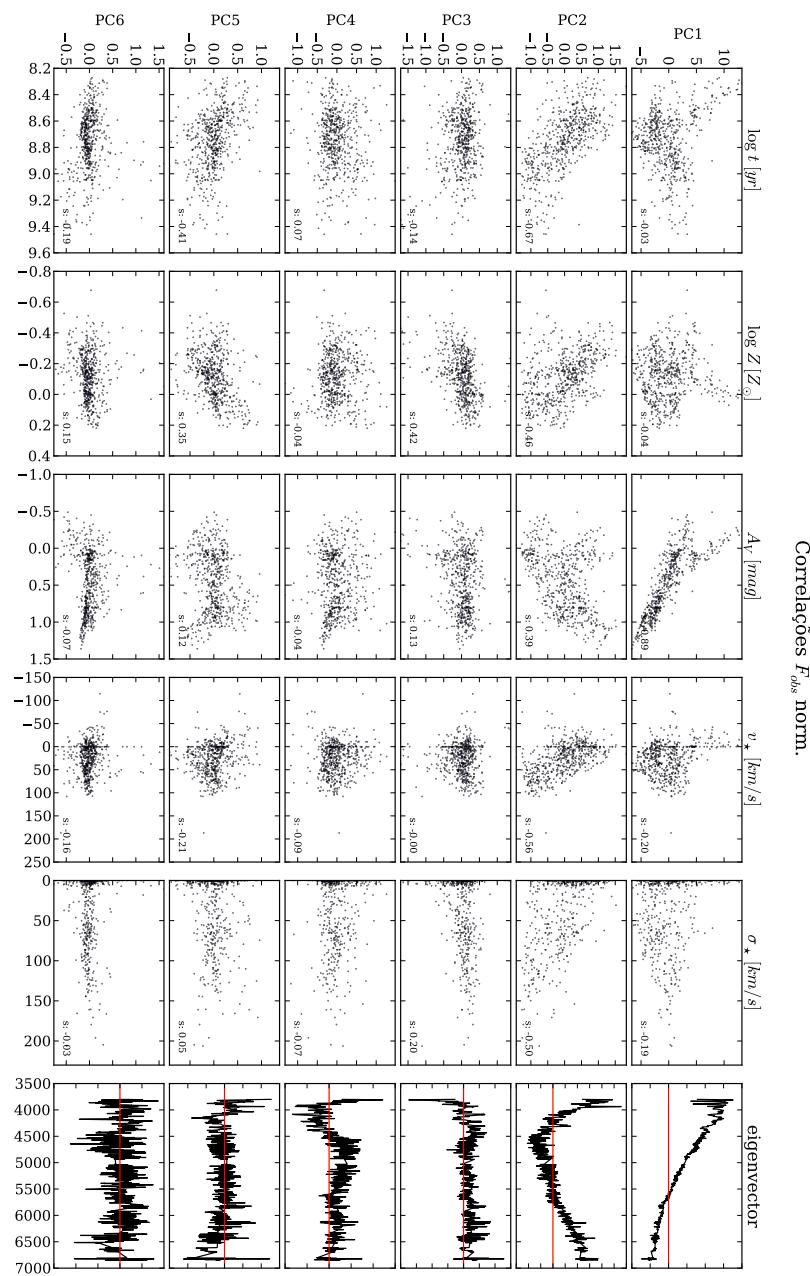


Figura 5.35: Igual a Figura 5.5 para a galáxia NGC 2623.

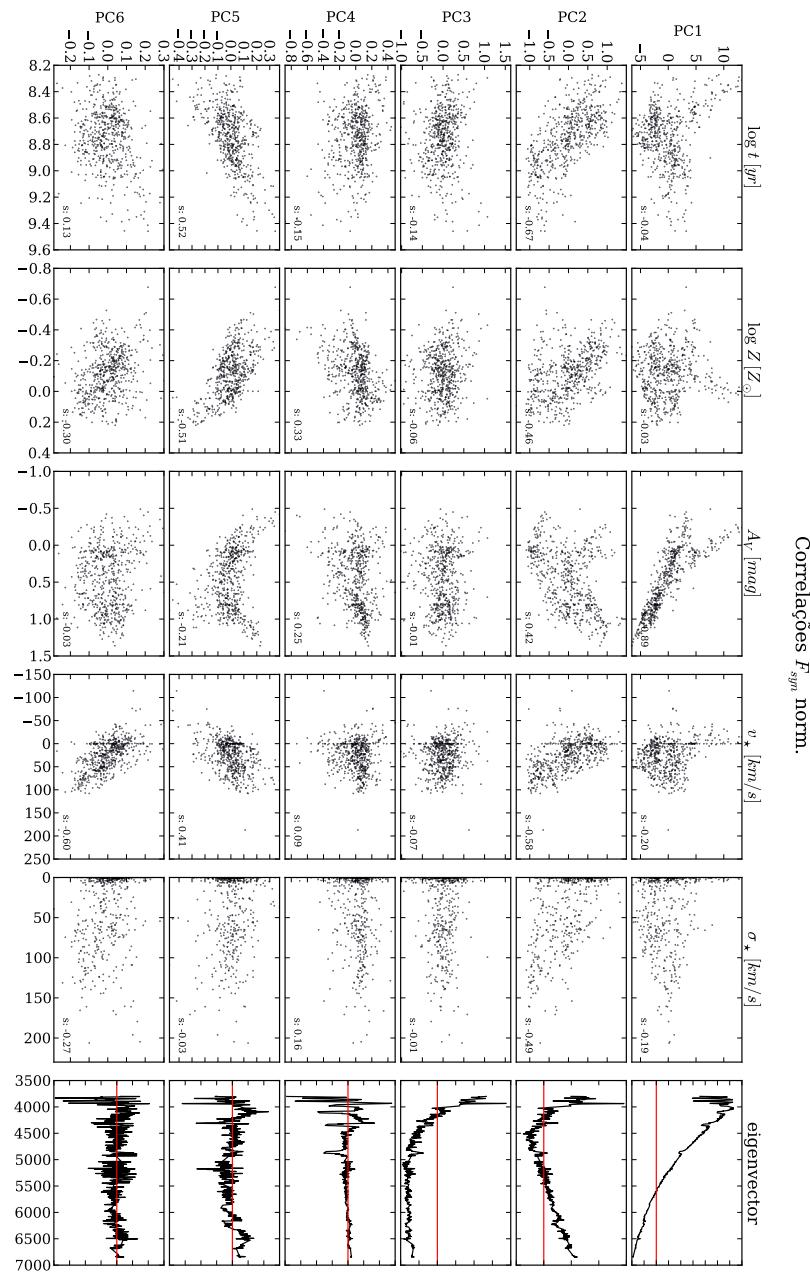
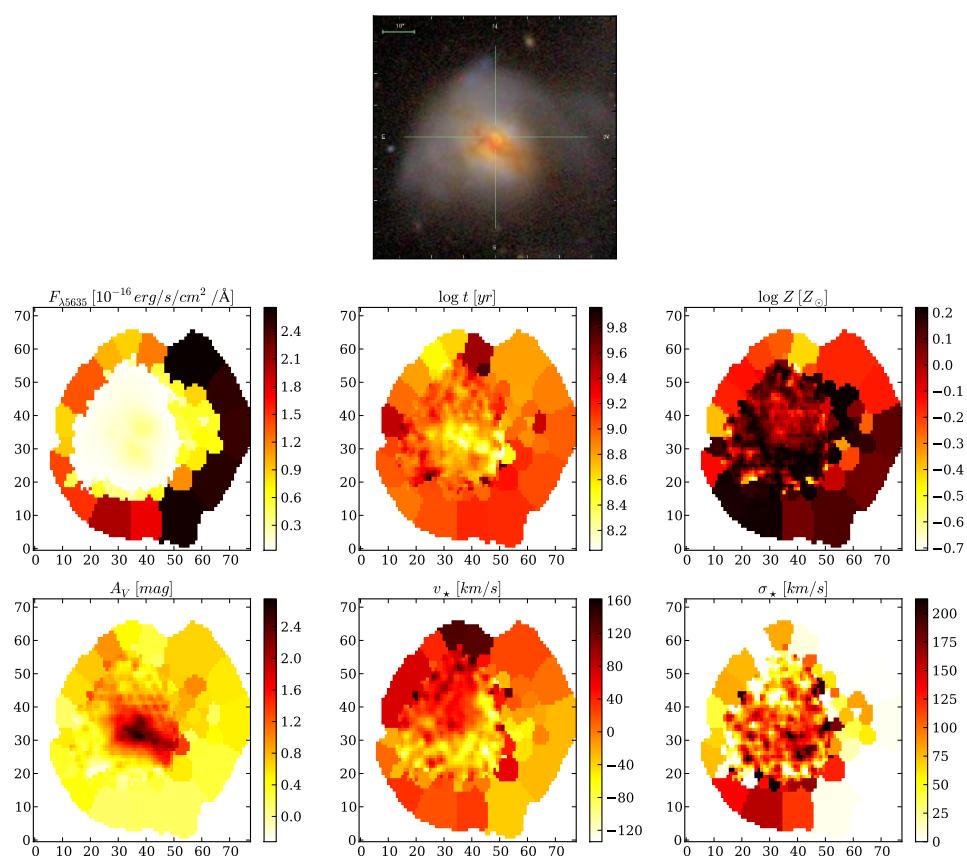


Figura 5.36: Igual a Figura 5.6 para a galáxia NGC 2623.

ARP220 - K0802

**Figura 5.37:** Igual a Figura 5.1 para a galáxia ARP 220.

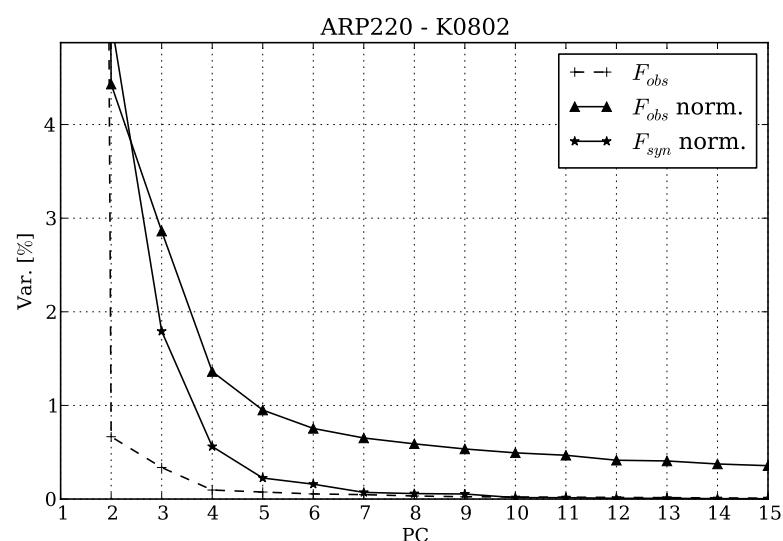


Figura 5.38: Igual a Figura ?? para a galáxia ARP 220.

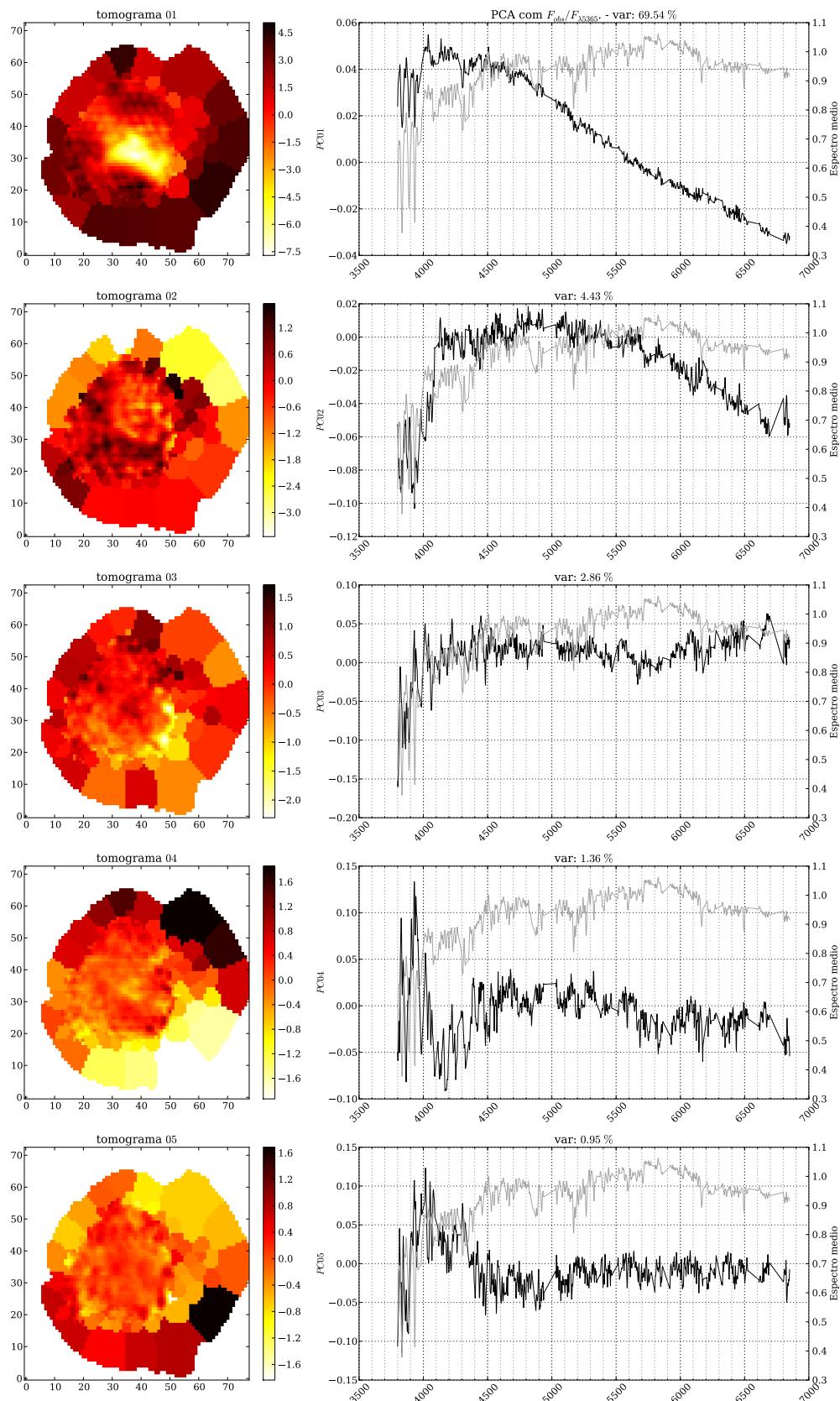


Figura 5.39: Igual a Figura 5.3 para a galáxia ARP 220.

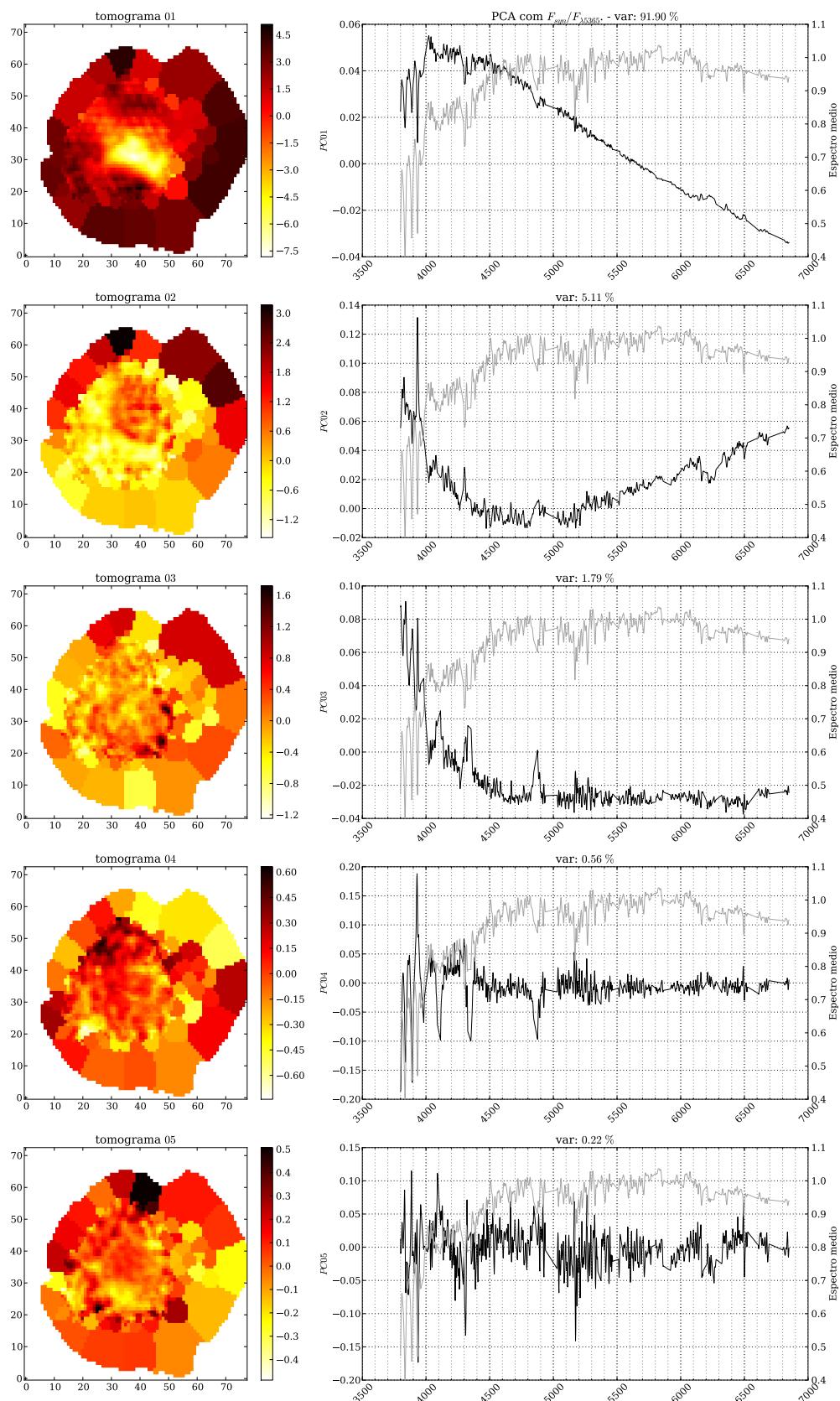


Figura 5.40: Igual a Figura 5.4 para a galáxia ARP 220.

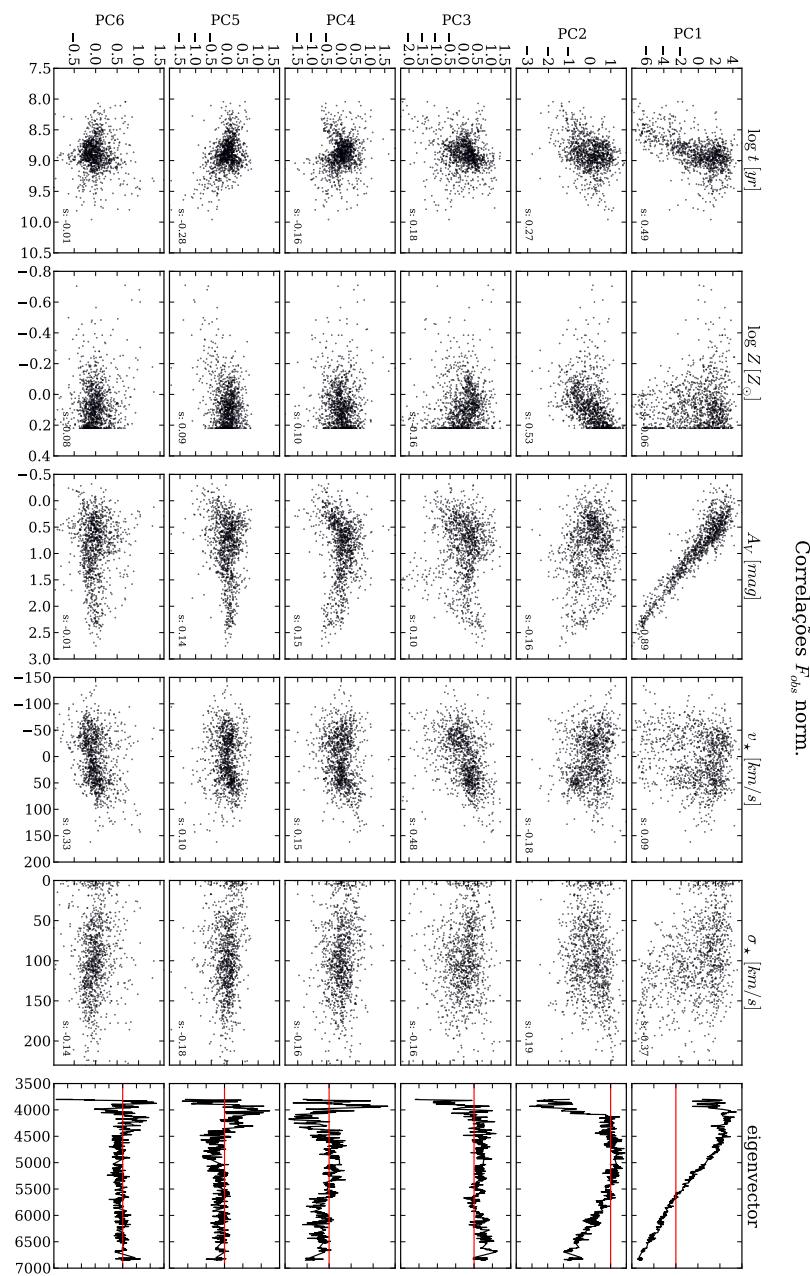


Figura 5.41: Igual a Figura 5.5 para a galáxia ARP 220.

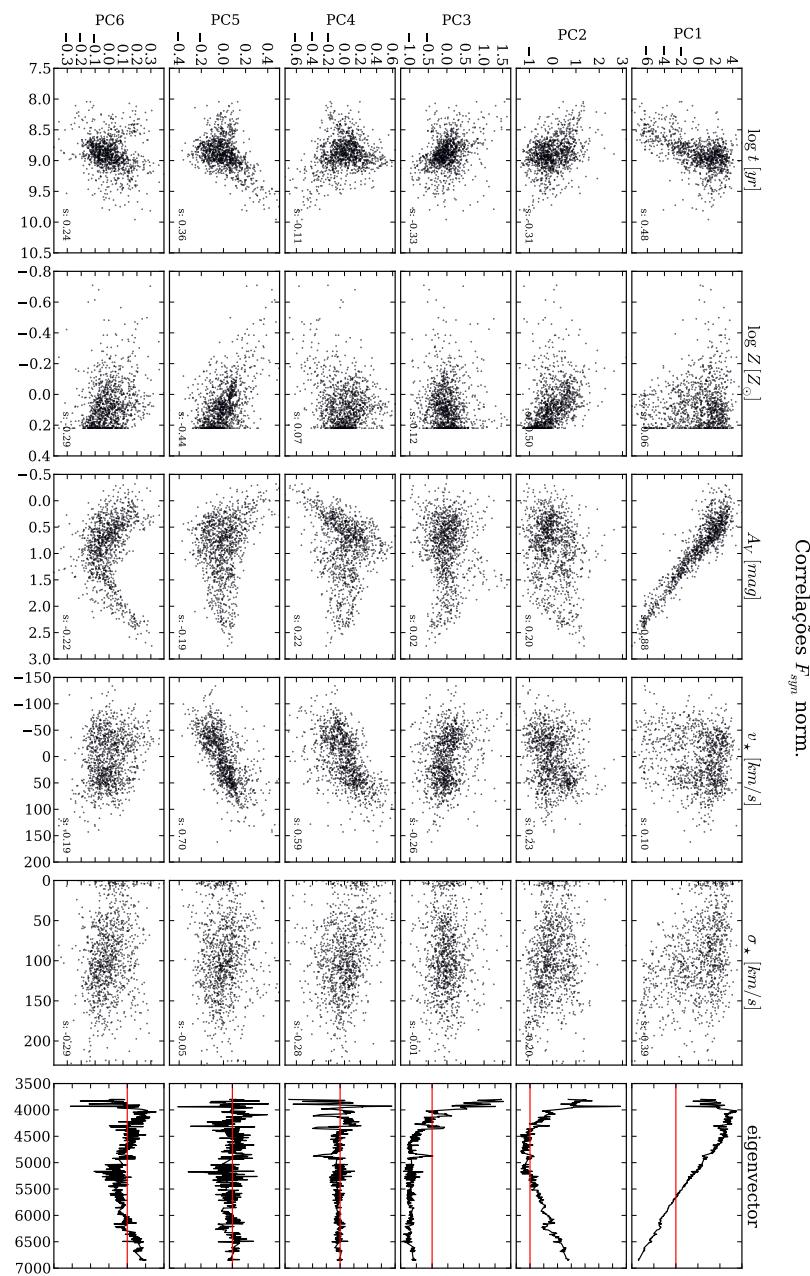


Figura 5.42: Igual a Figura 5.6 para a galáxia ARP 220.

Capítulo 6

Conclusões e perspectivas

6.1 Este trabalho

!○j○!

6.2 Trabalhos futuros

6.2.1 Linhas de emissão e intervalos específicos em comprimento de onda

Podemos também executar o PCA apenas em intervalos específicos do espectro. Isso pode ser feito de várias formas, por exemplo utilizando todos os pontos do(s) intervalo(s) em λ , ou utilizando apenas o fluxo integrado, ou apenas larguras equivalentes de linhas, ou FWHM das linhas. Um exemplo aplicado é fazer o PCA apenas das regiões que abrangem [O III]/H β em conjunto com [N II]/H α ou então H δ e D4000 para estudar a variância espacial desses ratios.

!○j○! Remover cinemática (v_0 & v_d) pois é um desperdício de variância. Incluir linhas de emissão na análise (aqui posso incluir umas figuras com essa parte que já está programada).

Apêndice A

[FIXME] Python

!○j○!

Referências Bibliográficas

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., Allende Prieto, C., An, D., Anderson, K. S. J., Anderson, S. F. et al. 2009, ApJS, 182, 543
- Ahn, C. P., Alexandroff, R., Allende Prieto, C., Anderson, S. F., Anderton, T., Andrews, B. H., Aubourg, É., Bailey, S. et al. 2012, ApJS, 203, 21
- Balakrishnan, G., Durand, F., & Guttag, J. 2013, 2013 IEEE Conference on Computer Vision and Pattern Recognition, 0, 3430
- Benítez, N., Gaztañaga, E., Miquel, R., Castander, F., Moles, M., Crocce, M., Fernández-Soto, A., Fosalba, P. et al. 2009, ApJ, 691, 241
- Blanc, G. A., Gebhardt, K., Heiderman, A., Evans, II, N. J., Jogee, S., van den Bosch, R., Marinova, I., Weinzierl, T. et al. 2010, in Astronomical Society of the Pacific Conference Series, Vol. 432, New Horizons in Astronomy: Frank N. Bash Symposium 2009, ed. L. M. Stanford, J. D. Green, L. Hao, & Y. Mao, 180
- Borcea, L., Callaghan, T., & Papanicolaou, G. 2012, CoRR, abs/1208.3700
- Chen, Y.-M., Kauffmann, G., Tremonti, C. A., White, S., Heckman, T. M., Kovač, K., Bundy, K., Chisholm, J. et al. 2012, MNRAS, 421, 314
- Cid Fernandes, R., González Delgado, R. M., García Benito, R., Pérez, E., de Amorim, A. L., Sánchez, S. F., Husemann, B., Falcón Barroso, J. et al. 2014, A&A, 561, A130
- Cid Fernandes, R., González Delgado, R. M., Schmitt, H., Storchi-Bergmann, T., Martins, L. P., Pérez, E., Heckman, T., Leitherer, C. et al. 2004, ApJ, 605, 105
- Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska, G., & Gomes, J. M. 2005, MNRAS, 358, 363
- Cid Fernandes, R., Pérez, E., García Benito, R., González Delgado, R. M., de Amorim, A. L., Sánchez, S. F., Husemann, B., Falcón Barroso, J. et al. 2013, A&A, 557, A86
- Colless, M. 1999, in Large-Scale Structure in the Universe, ed. G. Efstathiou & et al., 105
- Croom, S. M., Lawrence, J. S., Bland-Hawthorn, J., Bryant, J. J., Fogarty, L., Richards, S., Goodwin, M., Farrell, T. et al. 2012, MNRAS, 421, 872

- da Costa, L. N., Pellegrini, P. S., Sargent, W. L. W., Tonry, J., Davis, M., Meiksin, A., Latham, D. W., Menzies, J. W. et al. 1988, *ApJ*, 327, 544
- de Zeeuw, P. T., Bureau, M., Emsellem, E., Bacon, R., Carollo, C. M., Copin, Y., Davies, R. L., Kuntschner, H. et al. 2002, *MNRAS*, 329, 513
- González Delgado, R. M., Pérez, E., Cid Fernandes, R., García-Benito, R., de Amorim, A. L., Sánchez, S. F., Husemann, B., Cortijo-Ferrero, C. et al. 2013, ArXiv e-prints
- Huchra, J., Davis, M., Latham, D., & Tonry, J. 1983, *ApJS*, 52, 89
- Huchra, J. P. 1988, in Astronomical Society of the Pacific Conference Series, Vol. 5, The Minnesota lectures on Clusters of Galaxies and Large-Scale Structure, ed. J. M. Dickey, 41–70
- Husemann, B., Jahnke, K., Sánchez, S. F., Barrado, D., Bekerait*error*é, S., Bomans, D. J., Castillo-Morales, A., Catalán-Torrecilla, C. et al. 2013, *A&A*, 549, A87
- Iglesias-Páramo, J., Vílchez, J. M., Galbany, L., Sánchez, S. F., Rosales-Ortega, F. F., Mast, D., García-Benito, R., Husemann, B. et al. 2013, *A&A*, 553, L7
- Ivezic, Z., Tyson, J. A., Acosta, E., Allsman, R., Anderson, S. F., Andrew, J., Angel, R., Axelrod, T. et al. 2008, ArXiv e-prints
- Jolliffe, I. 2002, Principal Component Analysis, 2nd edn., Springer series in statistics (Springer)
- Kamruzzaman, S. M., Siddiqi, F. A., Islam, M. S., Haque, M. E., & Alam, M. S. 2010, CoRR, abs/1009.4974
- Kelz, A., Verheijen, M. A. W., Roth, M. M., Bauer, S. M., Becker, T., Paschke, J., Popow, E., Sánchez, S. F. et al. 2006, *PASP*, 118, 129
- Mateus, A., Sodré, L., Cid Fernandes, R., & Stasińska, G. 2007, *MNRAS*, 374, 1457
- Pérez, E., Cid Fernandes, R., González Delgado, R. M., García-Benito, R., Sánchez, S. F., Husemann, B., Mast, D., Rodón, J. R. et al. 2013, *ApJ*, 764, L1
- Ricci, T. V., Steiner, J. E., & Menezes, R. B. 2011, *ApJ*, 734, L10
- Riffel, R., Riffel, R. A., Ferrari, F., & Storchi-Bergmann, T. 2011, *MNRAS*, 416, 493
- Rosales-Ortega, F. F., Kennicutt, R. C., Sánchez, S. F., Díaz, A. I., Pasquali, A., Johnson, B. D., & Hao, C. N. 2010, *MNRAS*, 405, 735
- Roth, M. M., Kelz, A., Fechner, T., Hahn, T., Bauer, S.-M., Becker, T., Böhm, P., Christensen, L. et al. 2005, *PASP*, 117, 620
- Sánchez, S. F., Kennicutt, R. C., Gil de Paz, A., van de Ven, G., Vílchez, J. M., Wisotzki, L., Walcher, C. J., Mast, D. et al. 2012, *A&A*, 538, A8

- Sánchez Almeida, J., Aguerri, J. A. L., Muñoz-Tuñón, C., & de Vicente, A. 2010, ApJ, 714, 487
- Skrutskie, M. F., Cutri, R. M., Stiening, R., Weinberg, M. D., Schneider, S., Carpenter, J. M., Beichman, C., Capps, R. et al. 2006, AJ, 131, 1163
- Steiner, J. E., Menezes, R. B., Ricci, T. V., & Oliveira, A. S. 2009, MNRAS, 395, 64
- Trager, S. C., Faber, S. M., Worthey, G., & González, J. J. 2000, AJ, 119, 1645
- Verheijen, M. A. W., Bershady, M. A., Andersen, D. R., Swaters, R. A., Westfall, K., Kelz, A., & Roth, M. M. 2004, Astronomische Nachrichten, 325, 151
- York, D. G., Adelman, J., Anderson, Jr., J. E., Anderson, S. F., Annis, J., Bahcall, N. A., Bakken, J. A., Barkhouse, R. et al. 2000, AJ, 120, 1579