

Tomografia PCA aplicada 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
Programa 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) – 15 de março de 2014

A meus avós, pais e irmãos.

Agradecimentos

Difícil reunir agradecimentos para todas as contribuições de tanta gente boa que conheci durante esses últimos dois anos. Fácil mesmo é saber que sem minha família nada disso seria possível. Mãe e avós (Leila, Lourdes e Neide), sem vocês nada haveria. Seria impossível. Meus irmãos, Victória e José, sempre presentes. Tios e primos também sempre me dando força. Mesmo longe, estavam perto. Um agradecimento em especial ao meu pai *in memorian*, obrigado por me apontar a direção!

Quando, há três anos atrás, o meu grande amigo Prof. Kahio T. Mazon me indicou para meu orientador, Prof. Roberto Cid Fernandes Jr., eu não sabia o quanto ia ser legal. Aliás, tanto o Cid quanto o Kahio foram verdadeiros pais durante esses últimos anos. Obrigado a vocês dois que sempre souberam quando e como me ensinar e nunca deixaram proporcionar momentos que me incentivavam a seguir nessa caminhada!

Falando em ensinar e incentivar, agradeço também ao casal de amigos e professores Sônia e Fred, pelas horas de papo em corredores e salas de aulas da UFSC, me ajudando a enxergar as dificuldades e belezas das carreiras de professor e pesquisador.

Ao Prof. João Steiner e seus colaboradores pelo trabalho seminal desenvolvido. Ao Prof. Fabrício Ferrari pela sua participação na etapa inicial (e agora na final também como membro da banca examinadora). Esse trabalho também não seria possível sem o projeto CALIFA e todos os seus colaboradores. Muito obrigado a vocês. Em especial a Prof. Rosa M. Gonzalez Delgado e aos Profs. Sebastian F. Sánchez e Enrique Pérez, que são parte basilar para o andamento do projeto. Agradeço também ao amigo e pesquisador Rubén García Benito.

Durante esses dois anos pude viajar por muitos lugares e não passei em nenhum deles sem deixar amigos. Obrigado a todos aqueles que conheci através da astrofísica, seja pelas reuniões da SAB, ou pelas viagens em congressos, workshops e encontros, no Rio Grande do Sul, em São Paulo, na Venezuela ou no México. Em especial a Prof. Grażyna Stasińska e ao Prof. Christophe Morisset - um dos melhores fotógrafos que já conheci - que já se tornaram

mais do que amigos. Obrigado também aos amigos Alfredo Mejía, Glória Delgado e Ivan Cabrera! Estão sempre convidados a voltar ao Brasil.

Aos amigos da universidade. Em especial aos professores do GAS-UFSC: Abílio Mateus, Antônio Kanaan, Natalia Asari e Raymundo Baptista. Os amigos astrofísicos André, William, Bernardo, Marielli, Herpich, Germano, Rafael e Ariel; meus amigos e irmãos de casa, Maykot e Hari, além dos que nos acompanham nas insalubridades da vida, mesmo que por um curto, mas intenso, tempo: André 1 e 2, Bruno, Suasane e Ubiratã, Jéssica, Roger, Mateus “pinguim”, Adriano “PCM”, Diego e Vanessa e todos aqueles que de alguma forma me apoaram durante este tempo. Quero agradecer também aos professores e funcionários do Programa de Pós-Graduação em Física – especialmente ao Chefe de Expediente do PPGFSC, Antônio Machado.

Não poderia deixar de aqui agradecer aos amigos que estou em falta (e que também fazem muita falta). Acredito que minha vida será sempre melhor com a compreensão, amizade e amor de vocês, meus compadres, Fernando e Vivian, os amigos da música Leonardo e Adriana, Márcio e Elisabeth, Felipe e Sabrina, e ao meu irmão escolhido, Marcus Vinicius. Obrigado também a todas as suas famílias que ao longo dos anos já são parte da minha.

Obrigado a todos vocês!

Resumo

Levantamentos espectroscópicos de galáxias, como o *SDSS*, estão sendo levados para um novo nível através das unidades de campo integral (IFU), transformando o foco das propriedades das galáxias como um todo para a física interna das galáxias. O *survey* CALIFA é pioneiro nesta nova geração, proporcionando espectroscopia espacialmente resolvida para centenas de galáxias de todas as formas e massas. Vários estudos têm sido realizados com os cubos de dados deste *survey*, examinando as distribuições espaciais de linhas de emissão, populações estelares, cinemática do gás e das estrelas, etc. Neste trabalho exploramos a análise dos cubos do CALIFA através da perspectiva matemática da técnica de tomografia da análise de componentes principais (PCA) criada por Steiner e colaboradores. Utilizamos a tomografia PCA em 8 galáxias do CALIFA (4 *late-types*, 2 *early-types* e 2 *mergers*). Linhas de emissão são mascaradas em nossas análises para que nos concentremos nas propriedades de populações estelares.

As escalas físicas e angulares abarcadas pelos nossos cubos de dados (cobrindo galáxias inteiras) são muito maiores do que aquelas cobertas por aplicações anteriores desta técnica. Essa diferença nos impeliu a introduzir uma simples variação no estágio pré-PCA, normalizando os espectros a uma escala comum de fluxo, destacando variância espectral qualitativa (em oposição à quantitativa relacionada com a amplitude). Examinamos também cubos sintéticos obtidos com ajustes espectrais dos dados originais. Isto elimina o ruído e efeitos de calibração de fluxo. É bem sabido que não podemos atribuir sentido físico os resultados da PCA. Para resolver este problema, utilizamos uma análise de populações estelares feitas com o *STARLIGHT* sobre os mesmos dados para obtermos, através de uma engenharia reversa, o significado astrofísico das componentes principais.

Em cada galáxia fomos capazes de encontrar alguma correlação entre componentes principais e parâmetros físicos. Nós detectamos que a primeira componente principal em galáxias espirais geralmente se correlaciona com a idade estelar média ($\langle \log t \rangle_L$). Em mergers a componente principal reflete essencialmente variações da extinção global (A_V). Em alguns casos a tomografia PCA nos ajuda a localizar erros no cubos originais pois normalmente a grande variância imposta por pixels danificados aparece em componentes principais isolados.

Abstract

Spectroscopic surveys of galaxies, like the SDSS, are currently being taken to a next level with integral field units, turning the focus from the properties of galaxies as a whole to the internal physics of galaxies. The CALIFA survey is a pioneer in this new generation, providing spatially resolved spectroscopy for hundreds of galaxies of all shapes and masses. Several studies have been conducted with datacubes from this survey, examining the spatial distributions of emission lines, stellar populations, gaseous and stellar kinematics, etc. In this exploratory work we approach the analysis of CALIFA datacubes from the mathematical perspective of principal component analysis (PCA) tomography techniques developed by Steiner and collaborators. We apply the PCA tomography to 8 CALIFA galaxies. (4 *late-type*, 2 *early-type* and 2 *mergers*). Emission lines are masked in our analysis to concentrate on the stellar population properties.

The physical and angular scales spaned by our datacubes (which cover whole galaxies) are much larger than those spanned by previous applications of this technique. This difference prompted us to introduce a simple variation on the pre-PCA stage, normalizing all spectra in a datacube to a common flux scale, which highlight qualitative (as opposed to amplitude-related quantitative) spectral variance in the datacube. This eliminates noise and flux-calibration effects on the data. A well known caveat with PCA is that it does not assign physical meaning to its results. To tackle this problem we use a STARLIGHT-based stellar population analysis of the same data to reverse-engineer the astrophysical meaning of the principal components.

In every galaxy we are able to find some correlation between principal components and physical parameters. We find that the first principal component in spiral galaxies usually correlates with the mean stellar age ($\langle \log t \rangle_L$). In *mergers* the first component reflects essencially variations of the global extinction parameter (A_V). In some cases, the PCA tomography helps us find errors in the original datacube because usually the huge variance imposed by damaged pixels appears in isolated principal components.

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 os <i>pipelines</i> QBICK e PyCASSO	8
2.1	O survey CALIFA	8
2.2	Os <i>pipelines</i> QBICK e PyCASSO	11
2.2.1	QBICK	11
2.2.2	PyCASSO	15
2.2.3	Exemplos de utilização do PyCASSO	16
3	PCA e Tomografia PCA	20
3.1	Principal Component Analisys	20
3.1.1	PCA das galáxias do CALIFA	21
3.2	Tomografia PCA	24
3.3	Exemplo de resultado da Tomografia PCA	24
3.3.1	Descoberta de linhas largas em um LINER	24

3.4	Diferenças com respeito ao trabalho anterior	27
4	Aplicação a cubos de dados do CALIFA	29
4.1	PCA dos espectros observados: Efeitos de amplitude	30
4.2	PCA dos espectros sintéticos	36
4.3	Comparando as PCs com o <i>STARLIGHT</i> : engenharia reversa	40
5	Resultados: Tomografia PCA para 8 galáxias do CALIFA	47
5.1	Apresentação das galáxias	47
5.2	Galáxias espirais	48
5.2.1	NGC 0001 - CALIFA 8	48
5.2.2	NGC 0776 - CALIFA 73	49
5.2.3	NGC 4210 - CALIFA 518	57
5.3	Galáxias <i>early-type</i>	68
5.3.1	NGC 1167 - CALIFA 119	68
5.3.2	NGC 6515 - CALIFA 864	68
5.4	<i>Mergers</i>	81
5.4.1	NGC 2623 - CALIFA 213	81
5.4.2	ARP 220 - CALIFA 802	81
5.5	Considerações gerais	94
6	Conclusões e perspectivas	96
6.1	Este trabalho	96
6.2	Trabalhos futuros	97

Listas 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	Máscara espacial, zonas de Voronoi e flags para a galáxia NGC 2916 (CALIFA 277).	14
2.5	Exemplo de programa utilizando PyCASSO.	16
2.6	Programa idade estelar média.	17
2.7	Mapa da idade estelar média da galáxia NGC 2916 (CALIFA 277).	17
2.8	Exemplo de programa para perfil radial.	18
2.9	Perfil radial da idade estelar média da galáxia NGC 2916 (CALIFA 277).	18
3.1	Exemplo de cálculo de PCA usando o PyCASSO e SciPy.	23
3.2	<i>Scree test</i> na galáxia NGC 4736.	25
3.3	Tomograma e autoespectro 1 da galáxia NGC 4736.	26
3.4	Tomograma e autoespectro 2 da galáxia NGC 4736.	26
3.5	Tomograma e autoespectro 3 da galáxia NGC 4736.	27
4.1	Propriedades físicas da galáxia NGC 2916 (CALIFA 277).	31

4.2	Máximo e mínimo de fluxo dos espectros da galáxia NGC 2916 (CALIFA 277).	32
4.3	Tomogramas de 1 a 5 da galáxia NGC 2916 - F_{obs} .	34
4.4	Tomogramas de 1 a 5 da galáxia NGC 2916 - f_{obs} .	35
4.5	Fluxos de normalização para cada zona da galáxia K0277.	36
4.6	Scree test comparativo entre 3 PCAs.	37
4.7	Tomogramas de 1 a 5 da galáxia NGC 2916 - f_{syn} .	39
4.8	Correlações PCs vs. parâmetros físicos - f_{obs} .	41
4.9	Correlações PCs vs. parâmetros físicos - f_{syn} .	42
4.10	Dados no espaço das PCs vs AV- f_{obs} .	43
4.11	Dados no espaço das PCs vs AV- f_{syn} .	44
4.12	Correlações PCs vs. parâmetros físicos - F_{obs} .	46
5.1	Propriedades físicas da galáxia NGC 0001.	50
5.2	Scree test comparativo entre 3 PCAs - NGC 0001.	51
5.3	Tomogramas de 1 a 5 para o cubo f_{obs} - NGC 0001.	52
5.4	Tomogramas de 1 a 5 para o cubo f_{syn} - NGC 0001.	53
5.5	Correlações PCs vs. parâmetros físicos - f_{obs} - NGC 0001	54
5.6	Correlações PCs vs. parâmetros físicos - f_{syn} - NGC 0001	55
5.7	Propriedades físicas da galáxia NGC 0776.	56
5.8	Scree test comparativo entre 3 PCAs - NGC 0776.	57
5.9	Tomogramas de 1 a 5 para o cubo f_{obs} - NGC 0776.	58
5.10	Tomogramas de 1 a 5 para o cubo f_{syn} - NGC 0776.	59
5.11	Correlações PCs vs. parâmetros físicos - f_{obs} - NGC 0001	60
5.12	Correlações PCs vs. parâmetros físicos - f_{syn} - NGC 0001	61
5.13	Propriedades físicas da galáxia NGC 4210.	62

5.14 Scree test comparativo entre 3 PCAs - NGC 4210.	63
5.15 Tomogramas de 1 a 5 para o cubo f_{obs} - NGC 4210.	64
5.16 Tomogramas de 1 a 5 para o cubo f_{syn} - NGC 4210.	65
5.17 Correlações PCs vs. parâmetros físicos - f_{obs} - NGC 4210.	66
5.18 Correlações PCs vs. parâmetros físicos -um padrão de velocidades f_{syn} - NGC 4210.	67
5.19 Propriedades físicas da galáxia NGC 1167.	69
5.20 Scree test comparativo entre 3 PCAs - NGC 1167.	70
5.21 Tomogramas de 1 a 5 para o cubo f_{obs} - NGC 1167.	71
5.22 Tomogramas de 1 a 5 para o cubo f_{syn} - NGC 1167.	72
5.23 Correlações PCs vs. parâmetros físicos - f_{obs} - NGC 1167.	73
5.24 Correlações PCs vs. parâmetros físicos - f_{syn} - NGC 1167.	74
5.25 Propriedades físicas da galáxia NGC 6515.	75
5.26 Scree test comparativo entre 3 PCAs - NGC 6515.	76
5.27 Tomogramas de 1 a 5 para o cubo f_{obs} - NGC 6515.	77
5.28 Tomogramas de 1 a 5 para o cubo f_{syn} - NGC 6515.	78
5.29 Correlações PCs vs. parâmetros físicos - f_{obs} - NGC 6515.	79
5.30 Correlações PCs vs. parâmetros físicos - f_{syn} - NGC 6515.	80
5.31 Propriedades físicas da galáxia NGC 2623.	82
5.32 Scree test comparativo entre 3 PCAs - NGC 2623.	83
5.33 Tomogramas de 1 a 5 para o cubo f_{obs} - NGC 2623.	84
5.34 Tomogramas de 1 a 5 para o cubo f_{syn} - NGC 2623.	85
5.35 Correlações PCs vs. parâmetros físicos - f_{obs} - NGC 2623.	86
5.36 Correlações PCs vs. parâmetros físicos - f_{syn} - NGC 2623.	87
5.37 Propriedades físicas da galáxia ARP 220.	88

5.38 Scree test comparativo entre 3 PCAs - ARP 220.	89
5.39 Tomogramas de 1 a 5 para o cubo f_{obs} - ARP 220.	90
5.40 Tomogramas de 1 a 5 para o cubo f_{obs} - ARP 220.	91
5.41 Correlações PCs vs. parâmetros físicos - f_{obs} - ARP 220.	92
5.42 Correlações PCs vs. parâmetros físicos - f_{syn} - ARP 220.	93

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 juntos, através de um número interminável de combinações, abarcam o combustível para o desenvolvimento do conhecimento da humanidade. Qualquer que 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. Na Astrofísica tais informações hoje são reunidas através de grandes projetos chamados *surveys*.

Um *survey* astronômico é um levantamento de informações ou mapeamento de regiões do céu utilizando telescópios e detetores. Primordialmente surgem com a inata curiosidade do homem de observar e registrar tudo à sua volta. O crescimento desses catálogos é produto direto da evolução dos equipamentos usados nestas investigações. No início da década de 80, ?) constroem um *redshift-survey* que coletou espectros para 2500 galaxias dando início a produção sistemática de dados na astronomia extragalática. Esse trabalho foi seguidos por outros (e.g., ??) e de lá para cá a quantidade (e a qualidade) de dados só aumentou. Com a criação dos *mega-surveys* (*SDSS*, *2dFGRS*, *2MASS*; ???), alguns já terminados, outros ainda começando ou por terminar (*LSST*, *JPAS*; ??), 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 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 ??), 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 (e.g., ???). O STARLIGHT, assim como várias outras ferramentas similares (???) 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. ?), 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 (e.g., ???), as propriedades são as medidas de fluxo para cada comprimento de onda: $F_{gl} = F_g(\lambda_l)$, com $l = 1 \dots M$ comprimentos de onda, 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 (??).

Podemos ver exemplos também em medicina (?). 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 (?). Outro exemplo de uso de PCA aparece no artigo IV do grupo SEAGal/STARLIGHT (?) 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 ?) 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 (?) e pelo *Baryon Oscillation Spectroscopic Survey* (BOSS; ?), atribuindo um sentido físico a cada PC. Os trabalhos de ??? exemplificam outras aplicações de PCA a espectros do SDSS.

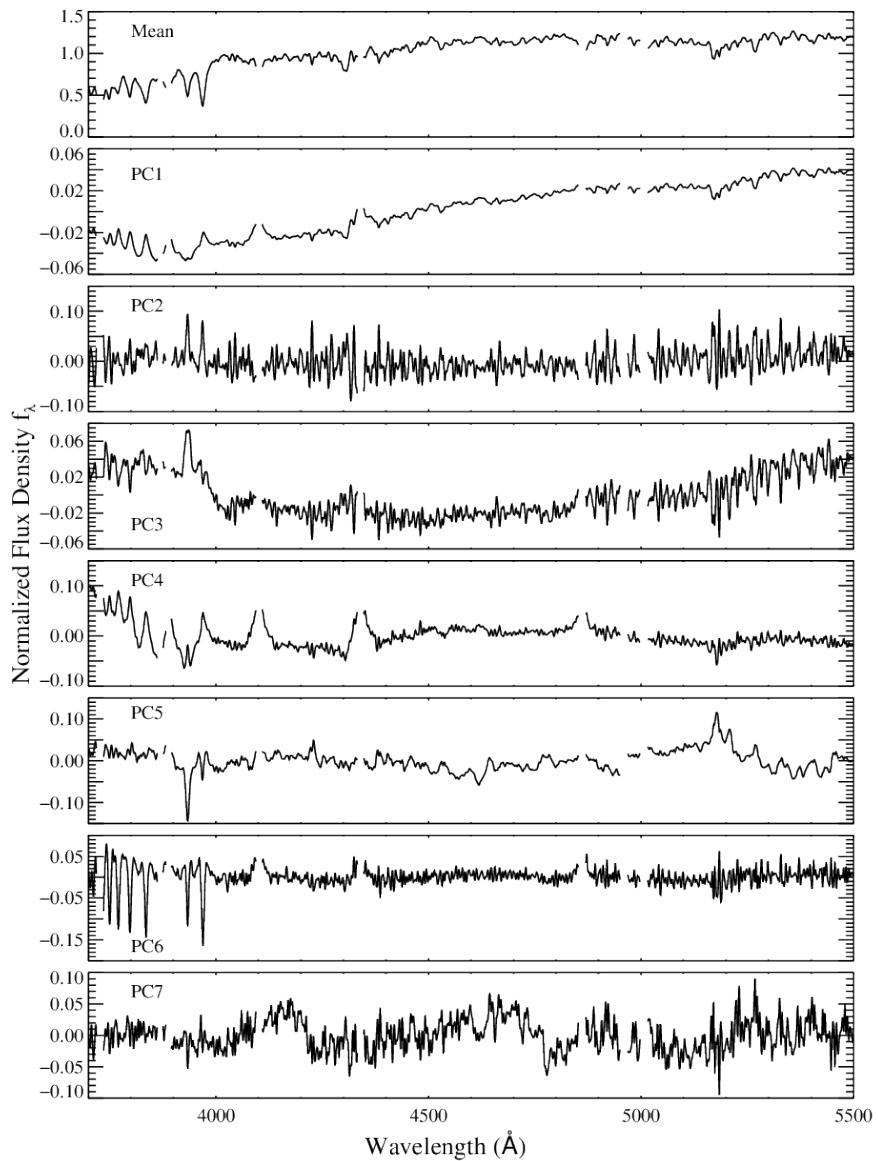


Figura 1.1: De cima para baixo: O espectro médio da biblioteca de espectros modelos seguido dos sete primeiros autoespectros da análise PCA. Retirado de ?).

1.2 A nova ferramenta - Tomografia PCA

Todos estes exemplos anteriormente citados relatam análises de espectros integrados de galáxias, i.e., espectros que contém toda luz (ou pelo menos boa parte) da galáxia. 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 ou variâncias). Existem também exemplos de análises PCA ou de síntese espectral baseadas em espectros nucleares (e.g., ??). 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* (*field-of-view*¹) de toda uma galáxia. 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; ?), 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 Instituto de Astrofísica de Andalucía, na Espanha (IAA; Enrique Pérez, Rosa González Delgado, Rubén García-Benito) e do Grupo de Astrofísica da Universidade Federal de Santa Catarina (GAS-UFSC; André Luiz de Amorim, Natália Vale Asari, Roberto Cid Fernandes). Aspectos técnicos e incertezas são discutidos em ?, CF13 daqui pra frente) e ?, enquanto ?, e ?) apresentam resultados astrofísicos dessa análise. Neste trabalho exploraremos os mesmo dados através da técnica de PCA.

A aplicação de PCA a cubos de dados foi realizada pela primeira vez por ?, 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 o peso, ou relevância, de cada componente principal (daqui pra frente PC) relacionada a uma posição na galáxia,

¹Campo de visão.

²spectral pixels

³<http://www.caha.es/CALIFA/>

combinando as virtudes de espectroscopia com as de imageamento.

Além de S09, outros exemplos de uso de Tomografia PCA podem ser encontrados em [?\) e ?](#)). Cumpre salientar que todas essas aplicações dessa nova técnica são baseadas em dados que cobrem apenas a região um campo de $3'' \times 5''$ correspondendo algumas centenas de parsecs centrais. Esta é naturalmente uma região rica em fenomenologia e portanto interessante de ser estudada com IFS, como ilustram os próprios resultados dos artigos supracitados (ver também Seção [3.2](#)).

1.3 Este trabalho

Através da colaboração do 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 ([?, 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 estejam 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 [?\), 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 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

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

utilizado é de aproximadamente 1.3 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 H_{II} 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 à 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 feita 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 os *pipelines* QBICK e PyCASSO

Nesse cenário de crescimento exponencial na quantidade de dados alavancados por novas tecnologias surge o CALIFA: um projeto de IFS que está modificando nossa maneira de ver e pensar as galáxias no nosso universo. 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. Na primeira seção deste capítulo iremos apresentar o survey CALIFA, que possibilita que este trabalho seja pioneiro na execução de análise PCA em cubos de espectros abrangendo o campo de uma galáxia completa.

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 ($\sim 10^6$). 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 pesquisas científicas mais abrangentes (SAURON; ?, região central de 72 galáxias com $z < 0.01$) (PINGS; ?, algumas galáxias muito próximas (~ 10 Mpc) e o estudo atual de 70 (ULIRGs com $z < 0.26$) (VENGA; ?, 30 galáxias espirais) (ATLAS^{3D}; ?, 260 galáxias *early-type* próximas). 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 (?) e MaNGA¹

A amostra-mãe do projeto contém 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* 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; ?) no modo PPAK (??) 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. Cada uma das *science fibers* (331) possui $2.7''$ de diâmetro, formando um campo de visão hexagonal de $74'' \times 64''$ ($\sim 1.3 \text{ arcmin}^2$) que, através de uma técnica de três pontos de *dithering* torna possível a observação de 100% do campo.

Os dados são reduzidos utilizando o CALIFA Pipeline versão 1.3c, descrito em ?). Com as 331 *science fibers* e o processo de *dithering* em 3 posições, temos 993 espectros independentes por objeto observado. ?) estimam que a resolução espacial final é $\sim 3.7''$ (FWHM). 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, somado ao *seeing* do céu e o processo de observação e redução, implica que cada píxel do cubo final não seja completamente independente, ou seja, numa automática correlação entre os pixels vizinhos. Os espectros vêm em duas configurações de rede: a V500 cobrindo de ~ 3700 até 7000 \AA com resolução de $\sim 6 \text{ \AA}$ de largura à meia altura (FWHM) e a V1200 ($\sim 3650 - 4600 \text{ \AA}$ FWHM $\sim 2.3 \text{ \AA}$). 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

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

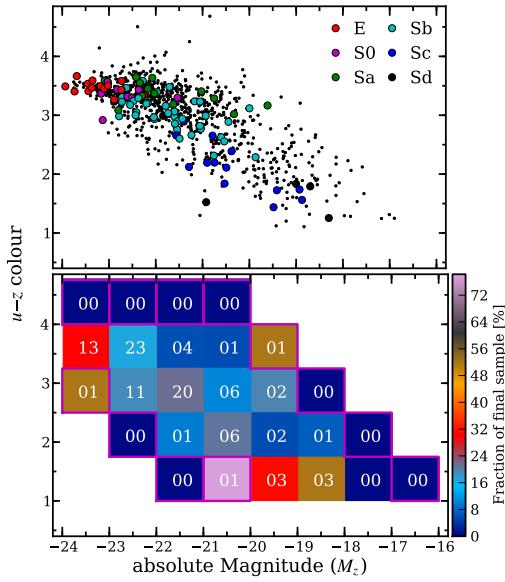


Figura 2.1: Distribuição das galáxias do CALIFA no diagrama cor magnitude $u-z$ vs. M_z . *Painel superior*: Em pontos pretos estão as galáxias pertencente a amostra-mãe e em cores as galáxias presentes no CALIFA DR1. As diferentes cores representam os diferentes tipos morfológicos. *Painel inferior*: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

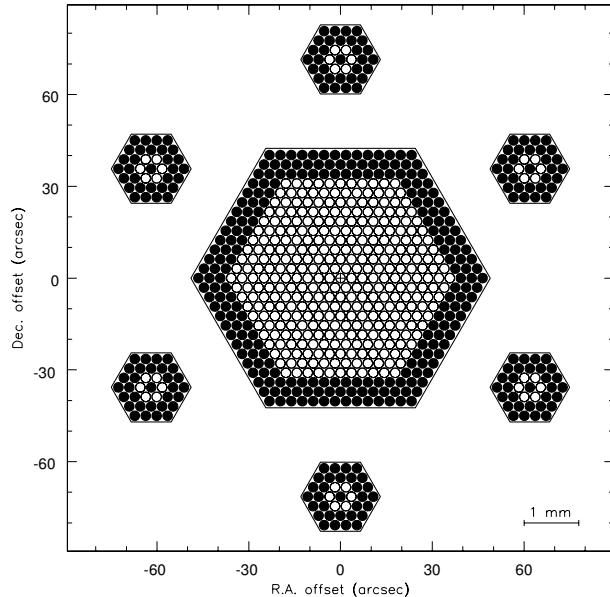


Figura 2.2: Este é o esquema com o *bundle* hexagonal com as 331 fibras de observação e mais 36 de amostra de céu. Retirado de ?).

redes, criando cubos que chamamos de COMBO: A parte com $\lambda < 4600 \text{ \AA}$ vem da V1200 e o resto da V500. Nessa montagem os espectros da V1200 foram suavizados à mesma resolução espectral da V500 ($\text{FWHM} = 6 \text{ \AA}$). Como outros cubos do CALIFA, os cubos COMBO provêm, para cada posição e comprimento de onda, o fluxo ($F_{\lambda,x,y}^{\text{orig}}$), seu erro ($\epsilon_{\lambda,x,y}$) e uma flag ($b_{\lambda,x,y}$) que sinaliza pixels defeituosos.

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

2.2 Os *pipelines* QBICK e PyCASSO

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

2.2.1 QBICK

O primeira é o QBICK, desenvolvido por Rubén García-Benito, do IAA, e explicado em detalhe em CF13. Resumidamente, este pipeline:

1. Cria máscaras espaciais que removem regiões externas de baixo S/N, e objetos indesejáveis no *FoV* (estrelas ou galáxias).
2. Introduz flags espetrais 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 \AA .
4. Define zonas de Voronoi, tal que o espectro total dessas zonas tem uma relação $S/N \geq 20$ em uma janela espectral de 90 \AA de largura e centrada em 5635 \AA , e, por fim, extrai os espectros correspondentes.

A Figura 2.3 ilustra as máscaras espetrais 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

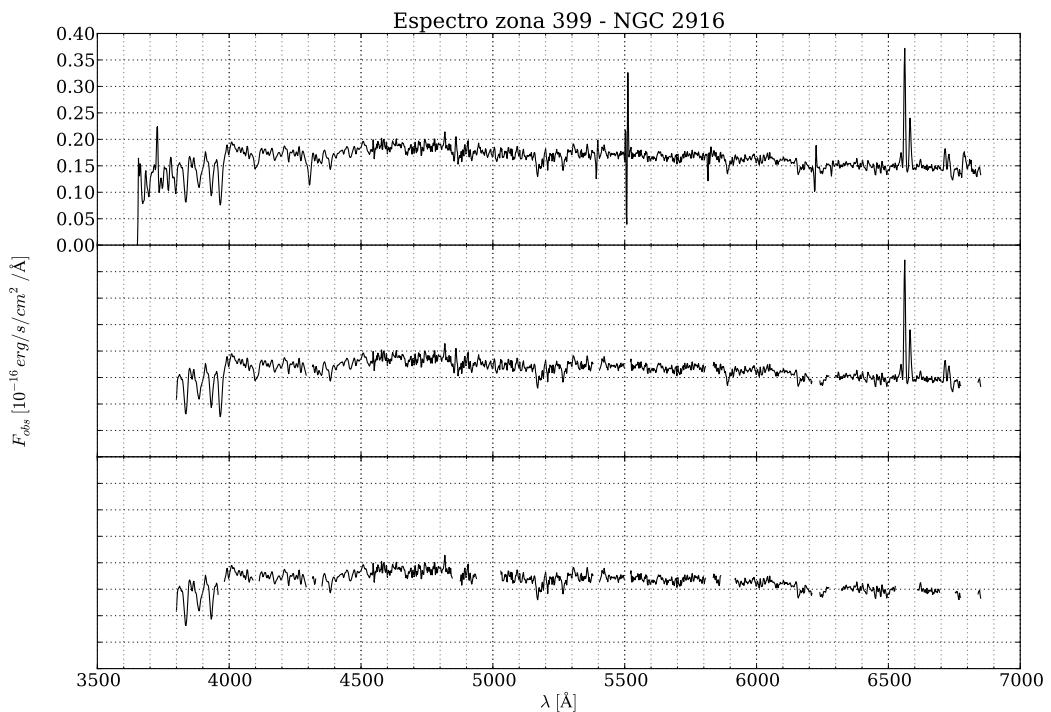


Figura 2.3: Espectro da zona 399 da galáxia NGC 2916 (CALIFA 277). Acima está o espectro completo. No segundo vemos o espectro com linhas telúricas e bad-pixels removidos, além do limite de intervalo em comprimento de onda de 3800 a 6850 Å. No espectro mais abaixo, além das partes removidas no segundo,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

correspondem a regiões contendo linhas telúricas², as quais dificilmente são perfeitamente removidas na redução. Pixels identificados como problemáticos durante a redução dos dados são 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 (ver Equação 3.5).

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 Na I D³: 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 são processados com o STARLIGHT. Também nessa etapa os espectros são cortados entre 3800 e 6850 Å. O espectro após 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 pixels úteis (i.e., não mascarados espacialmente) são agrupados em $N_z = 1638$ 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 píxel e ~ 3% com mais de 10 pixels agrupados, mesmo com a imposição de relação $S/N \geq 20$ na janela de 5635 ± 45 Å, nossa janela de comprimento de onda usada como referência. Podemos ver na Tabela 2.1 esses dados para as galáxias estudadas nesse trabalho. Portanto o cubo original é transformado numa matriz de zonas e comprimentos de onda: F_{zl} com $z = 1 \dots N_z$ e $l = 3800 \dots 6850$ Å.

A motivação para a “zonificação” do cubo é garantir espectros 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 razão para isso é que queremos comparar os resultados de nossa análise com os parâmetros extraídos pelo STARLIGHT, o que servirá de auxílio a interpretação dos resultados da PCA.

Todas informações e pré-processamentos provenientes do QBICK são herdadas pelo *pipeline* apresentado a seguir.

²Linhas de emissão ou absorção referentes à atmosfera.

³Na I D não é uma linha de emissão, mas mascaramos esse dbleto porque ele pode conter uma componente de absorção devida a presença de gás neutro no meio interestelar.

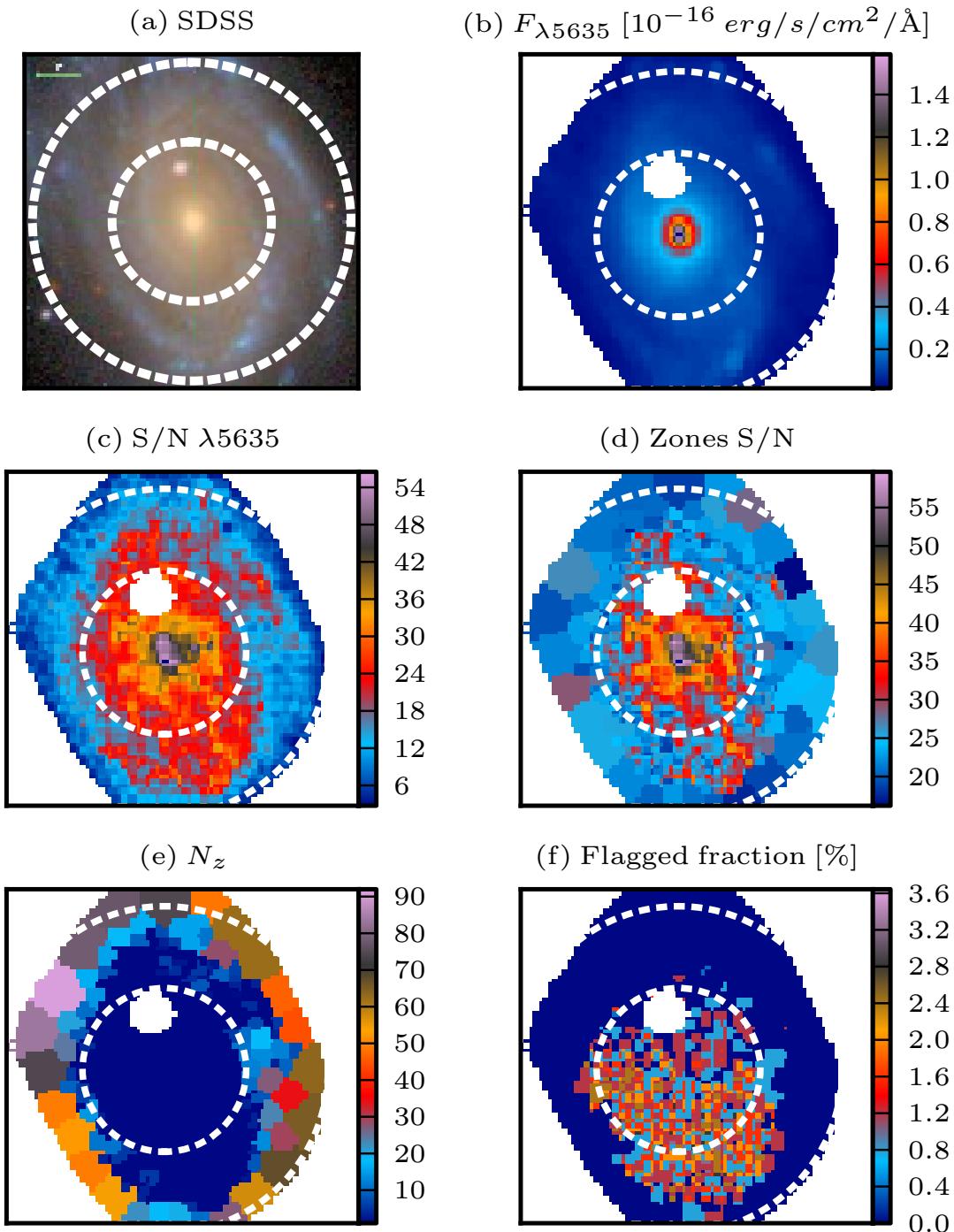


Figura 2.4: (a) Imagem SDSS para a galáxia NGC 2916 (CALIFA 277). (b) Fluxo no intervalo de 5635 ± 45 Å depois da aplicação da máscara espacial. (c) Mapa da relação S/N em 5635 Å. (d) Mapa da relação S/N depois da zonificação de Voronoi. (e) Número de spaxels em cada zona de Voronoi. (f) Percentagem de *bad pixels* no intervalo de 3800–6850 Å. Retirado de CF13.

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

2.2.2 PyCASSO

Os espectros extraídos pelo QBICK são processados pelo STARLIGHT, produzindo, no caso de CALIFA 277, 1639 arquivos ASCII como output. Dado o grande número de informações contidas em cada um desses arquivos (e.g., história de formação estelar em luz e massa, extinção, parâmetros cinemáticos, figuras de mérito do ajuste), necessitamos de um *pipeline* que faça a organização de todos os dados para que cálculos e gráficos 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*, descrito na sec. 4 de CF13) que faz a organização dos dados que vêm do QBICK e das etapas iniciais de redução 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 ex-

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.

ploratória 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. Todos os códigos desenvolvidos nessa dissertação foram escritos de forma compatível com o PyCASSO.

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 no programa da Figura 2.6 (e a imagem gerada na Figura 2.7) é feita através da função de `zoneToYX` dentro do PyCASSO. Os gráficos são feitos utilizando a biblioteca gráfica de Python, `matplotlib`⁵.

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

Já existem vários artigos utilizando o PyCASSO (????), também há alguns usando indire-

⁵<http://matplotlib.org>

```

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.

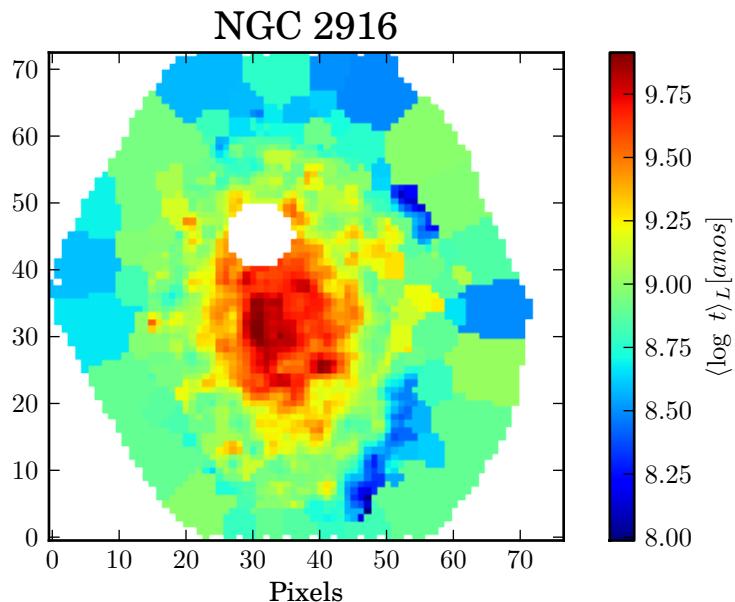


Figura 2.7: Mapa de idade estelar média ponderada pela luminosidade da galáxia NGC 2916 (CALIFA 277) gerado pelo programa da Figura 2.6.

```

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 LobsnSD__r = K.radialProfile(K.LobnSD_yx, bin_r=bins, rad_scale=1, mode='sum')
13 at_flux_LobsnSD__r = K.radialProfile(K.at_flux_yx * K.LobnSD_yx,
14                                         bin_r=bins, rad_scale=1, mode='sum')
15 at_flux__r = at_flux_LobsnSD__r / LobsnSD__r
16 # Desenhar o perfil.
17 import matplotlib.pyplot as plt
18 plt.xlabel('radius [arcsec]')
19 plt.ylabel(r'\langle \log t \rangle_L [anos]')
20 cb = plt.colorbar()
21 plt.plot(bin_center, at_rad)

```

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

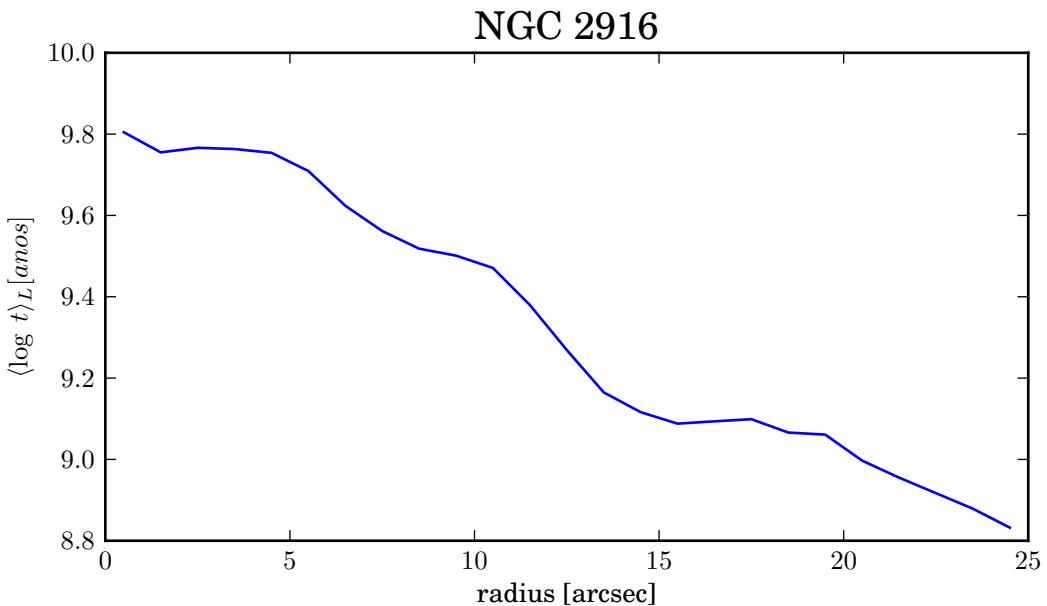


Figura 2.9: Perfil radial da idade estelar média ponderada pela luminosidade da galáxia NGC 2916 (CALIFA 277) gerado pelo programa da Figura 2.8.

tamente (??) e três teses de doutorado em curso (uma na UFSC e duas no IAA). São mais de 10 pessoas utilizando a biblioteca em nossa colaboração.

Os cálculos necessários para a PCA e também para a Tomografia PCA se tornam simples contas usando pacotes matemáticos de python. O software desenvolvido para a PCA e Tomografia PCA desta dissertação (por ora apelidado de “PCAlifa”) foi feito de forma compatível com o PyCASSO, visando futuramente integrá-lo como um novo módulo dessa ferramenta.

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 da PCA e de sua versão para cubos de dados (Tomografia PCA).

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) que maximizam o valor de $\mathbf{e}_k^T \cdot \mathbf{C}_{cov} \cdot \mathbf{e}_k$ (ver Eq. 3.5) 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 ($\mathbf{e}_i^T \mathbf{e}_i = 1$) (? , p. 5-6). No caso de PCA com espectros, o fluxo em cada comprimento de onda varia através das zonas. Encontramos então 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 covariância (a matriz tem

dimensão $N_\lambda \times N_\lambda$) usando a biblioteca científica SciPy¹.

3.1.1 PCA das galáxias do CALIFA

Conforme a Seção 2.2 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 ($N_\lambda \times N_z$) com N_λ comprimentos de onda e N_z zonas (f_obs no PyCASSO). Em nosso trabalho vamos usar a forma transposta dessa matriz, portanto nossa matriz de espectros está na forma $N_z \times N_\lambda$, como demonstrado na equação abaixo.

$$\mathbf{F}_{z\lambda} = \begin{bmatrix} F_{z_1\lambda_1} & F_{z_1\lambda_2} & F_{z_1\lambda_3} & \dots & F_{z_1\lambda_{N_\lambda}} \\ F_{z_2\lambda_1} & F_{z_2\lambda_2} & F_{z_2\lambda_3} & \dots & F_{z_2\lambda_{N_\lambda}} \\ F_{z_3\lambda_1} & F_{z_3\lambda_2} & F_{z_3\lambda_3} & \dots & F_{z_3\lambda_{N_\lambda}} \\ \dots & \dots & \dots & \dots & \dots \\ F_{z_{N_z}\lambda_1} & F_{z_{N_z}\lambda_2} & F_{z_{N_z}\lambda_3} & \dots & F_{z_{N_z}\lambda_{N_\lambda}} \end{bmatrix} \quad (3.1)$$

Calculamos então o espetro médio de uma galáxia

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

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:

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

Vemos que a matriz de covariância possui dimensão $N_\lambda \times N_\lambda$. Um elemento dessa matriz é

¹<http://scipy.org/>

dado por

$$\mathbf{C}_{ij} = \frac{\sum_{k=1}^{N_z} I_{z_k \lambda_i} I_{z_k \lambda_j}}{N_z - 1} \quad (3.5)$$

e expressa a covariância entre os fluxos nos comprimentos de onda λ_i e λ_j para todas as zonas. Os elementos da diagonal (\mathbf{C}_{ii}) expressam a variância dos fluxos no comprimento de onda λ_i entre todas as zonas.

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. Um exemplo de programa para calcular a matriz de covariâncias e seus autovetores e autovalores usando o PyCASSO e o SciPy pode ser visto na Figura 3.1.

Muitas figuras de PCs e suas utilizações e diferenças nos pré-processamentos serão mostradas nos Capítulos 4 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 armazenado nos vetores `eigval` e `eigvect` os autovalores e autovetores da matriz de covariância (`covMat_11`) ordenados decrescentemente.

3.2 Tomografia PCA

Os autoespectros (\mathbf{e}_k) da matriz de covariância ordenados pela sua variância (Λ_k) formam uma matriz ($\mathbf{E}_{\lambda k}$, de dimensão $N_\lambda \times N_k$, onde N_k é o número de PCs) que serve como base onde projetamos nossa matriz de observáveis com a média subtraída ($\mathbf{I}_{z\lambda}$) através da transformação:

$$\mathbf{T}_{zk} = \mathbf{I}_{z\lambda} \cdot \mathbf{E}_{\lambda k} \quad (3.6)$$

De posse dessa nova matriz transformada e de um mapa que faça a transformação 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 PCA², criada e assim batizada por S09, que fazem 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 (PC) relacionado que, em conjunto, trazem novas perspectivas e ideias para a interpretação de ambos.

3.3 Exemplo de resultado da Tomografia PCA

Para ilustrar o potencial da técnica de Tomografia PCA, ilustramos os resultados obtidos por S09.

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 Ionization 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*).?) 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 classificação desses objetos como tipo 1 ou 2. A PCA, juntamente com a Tomografia PCA, fazem esse papel da subtração das populações estelares sem haver nenhuma parametrização.

²<http://www.astro.iag.usp.br/~pcatomography/>

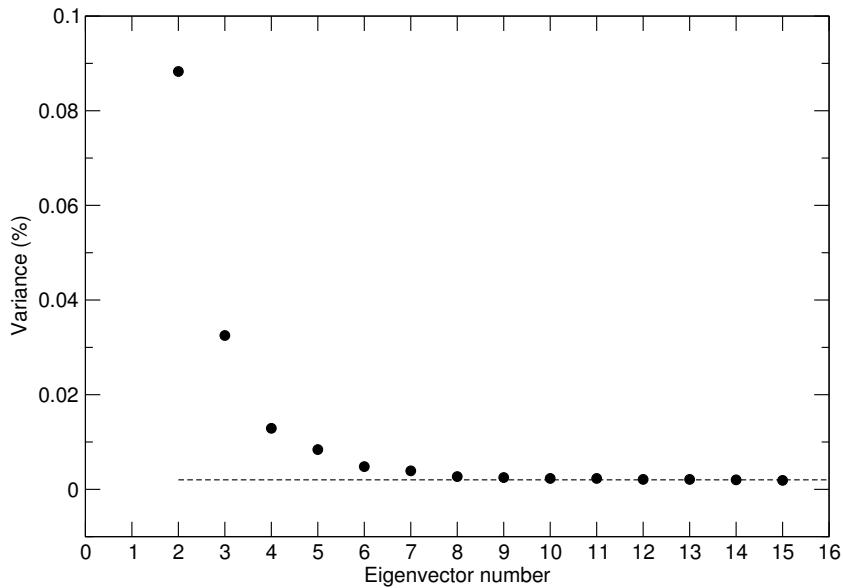


Figura 3.2: Scree test das primeiras 16 PCs do cubo de espectros da região central da galáxia NGC 4736. Retirado de S09.

S09 estudaram os 8 primeiros autoespectros, escolhidos através de um *scree test* (Figura 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 como 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 o 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.



Figura 3.3: Autoespectro 1 e seu respectivo tomograma. Retirado de S09.

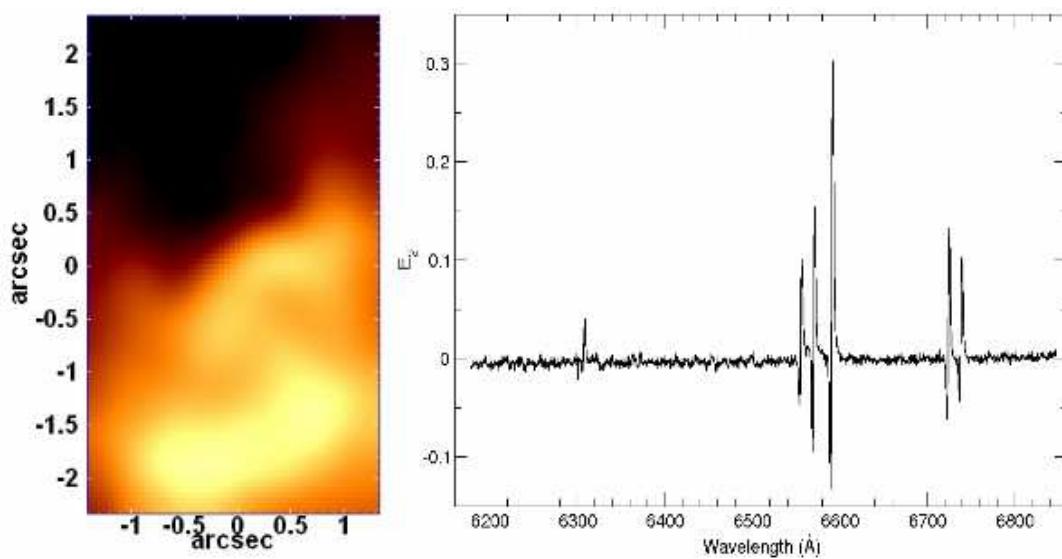


Figura 3.4: Autoespectro 2 e seu respectivo tomograma. Retirado de S09.

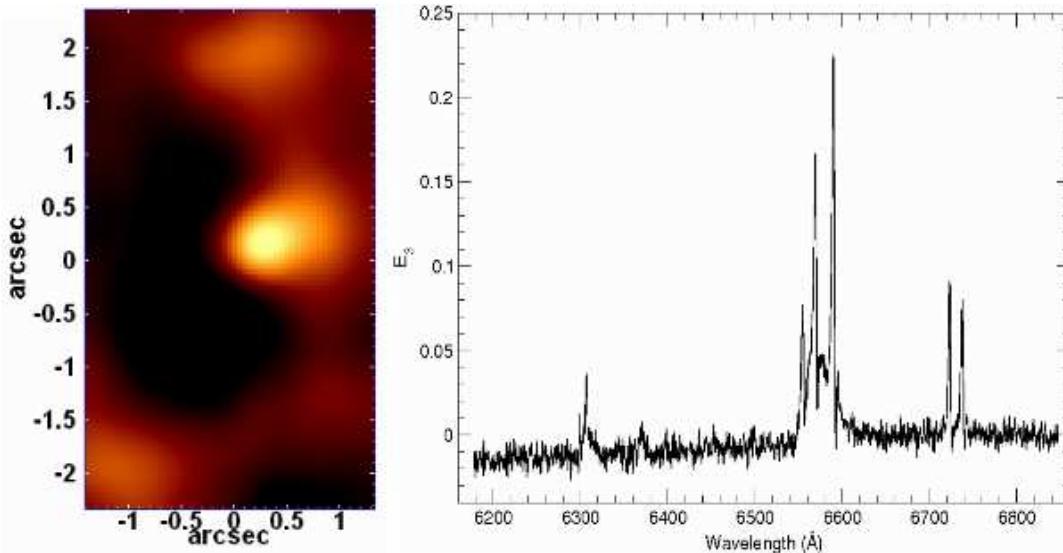


Figura 3.5: Autoespectro 3 e seu respectivo tomograma. Retirado de S09.

3.4 Diferenças com respeito ao trabalho anterior

O exemplo acima nos motiva a aplicar esta mesma técnica aos cubos de dados do CALIFA, e o resto desta dissertação apresenta nossos experimentos 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, S09 trabalham com dados do Gemini (8m) com 3×20 min de integração, enquanto nossos dados vêm de integrações de tipicamente 30 min para o V1200 e 15 min para o V500 com $2 \sim 3$ exposições, no telescópio de 3.5m do observatório de Calar Alto.

Além disso, eles 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. Não aplicaremos isso no nosso caso porque nosso *FoV* é muito maior, e não esperamos grandes ganhos em resolução espacial com a aplicação de tais técnicas de deconvolução espacial.

Como dito na Seção 2.1 a escala e resolução espacial de nossos dados difere brutalmente daquelas nos trabalhos acima resumidos: $3'' \times 5''$ e resolução de $0.47''$ no Gemini contra $74'' \times 64''$ e resolução de $\sim 3.7''$ no CALIFA. Portanto, o *FoV* estudado por S09 equivalem a aproximadamente ~ 2 elementos de resolução do CALIFA! Isto de cara mostra que a ciência

que podemos obter com Tomografia PCA aplicada aos dados do CALIFA não será a mesma que aquela estudada por S09.

Em resumo, apesar de inspirado diretamente pelos resultados de S09, este estudo difere muito dos trabalhos deles. Estas diferenças ficarão evidentes já a partir do próximo capítulo, onde apresentamos nossos primeiros resultados e discutimos algumas variações com respeito ao processamento de S09, que decidimos fazer para melhor se adequar ao contexto de dados do CALIFA.

Capítulo 4

Aplicação a cubos de dados do CALIFA

Com os dados e ferramentas apresentados nos capítulos anteriores estamos prontos para aplicar a Tomografia PCA a galáxias do CALIFA. Como a 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.

Todas as análises neste capítulo serão feitas usando a galáxia NGC 2916 (CALIFA 277), uma espiral de tipo Sbc (Figura 4.1). 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 Figura 4.1, enquanto a Figura 4.2 mostra o fluxo mínimo e máximo por comprimento de onda, dos espectros observados desse objeto. No painel direito vemos o mesmo mas para os espectros observados normalizados.

A Figura 4.3 apresenta os resultados da tomografia PCA para esse cubo de dados. O tomograma é formado pelo “peso” da PC em cada zona (ver Equação 3.6). 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 auxílio de seu tomograma vemos que para as zonas mais internas da galáxia (onde as populações são mais velhas, como vimos nas Figuras 2.7 e 4.1) 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 a PCA para o mesmo cubo, mas com os espectros normalizados, i.e., reescalando cada espectro de modo que o fluxo médio na janela de $5590 - 5680$ Å seja igual a 1. Daqui para frente vamos nos referir aos espectros normalizados utilizando a letra f :

$$f_{obs} = \frac{F_{obs}}{F_{\lambda 5635}} \quad (4.1)$$

A tomografia PCA desse cubo normalizado é mostrada na Figura 4.4. À primeira vista os resultados podem ser parecidos com o anterior deslocando uma componente acima, ou seja, a mesma 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

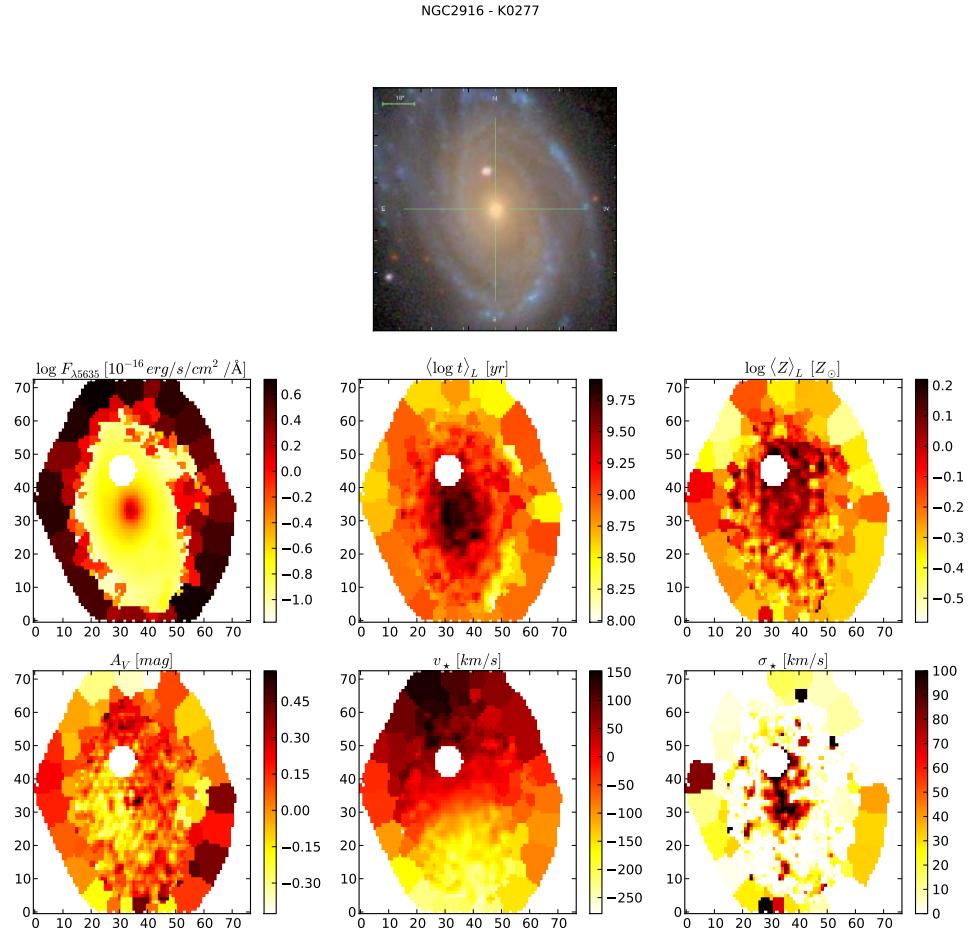


Figura 4.1: Propriedades físicas da galáxia NGC 2916 (CALIFA 277). Na primeira linha a imagem do SDSS da galáxia. Na primeira imagem da segunda linha temos o valor do fluxo observado em 5635 Å, usado para a normalização dos espectros de cada zona. Da segunda imagem da segunda linha em diante temos as propriedades físicas que serão correlacionadas com as PCs. São elas: $\langle \log t \rangle_L$, $\log \langle Z \rangle_L$, A_V , v_\star , σ_\star .

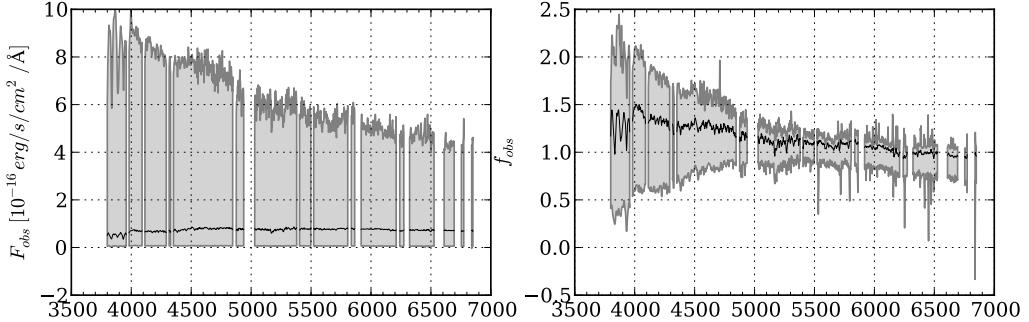


Figura 4.2: Fluxo máximo e mínimo por comprimento de onda para o cubo de espectros da galáxia NGC 2916 (CALIFA 277). O painel à esquerda é para os espectros observados (F_{obs}) e o à direita para os espectros observados e normalizados pelo fluxo mediano na janela de 5635 ± 45 Å (f_{obs}). Em destaque o 95º percentil. Todos os espectros do cubo já estão mascarados da mesma forma que o terceiro painel da Figura 2.3

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 claro padrão de rotação na PC2 da Figura 4.4. Com uma saturação na escala de cores no tomograma da PC3 da Figura 4.3 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. Uma PCA nessa galáxia hipotética não identificaria nenhuma fonte de variancia, pois todos 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 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 des-
cartável 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 4.3) e a imagem mais à esquerda na Figura 4.5 formada pelos fluxo 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 participam com 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.

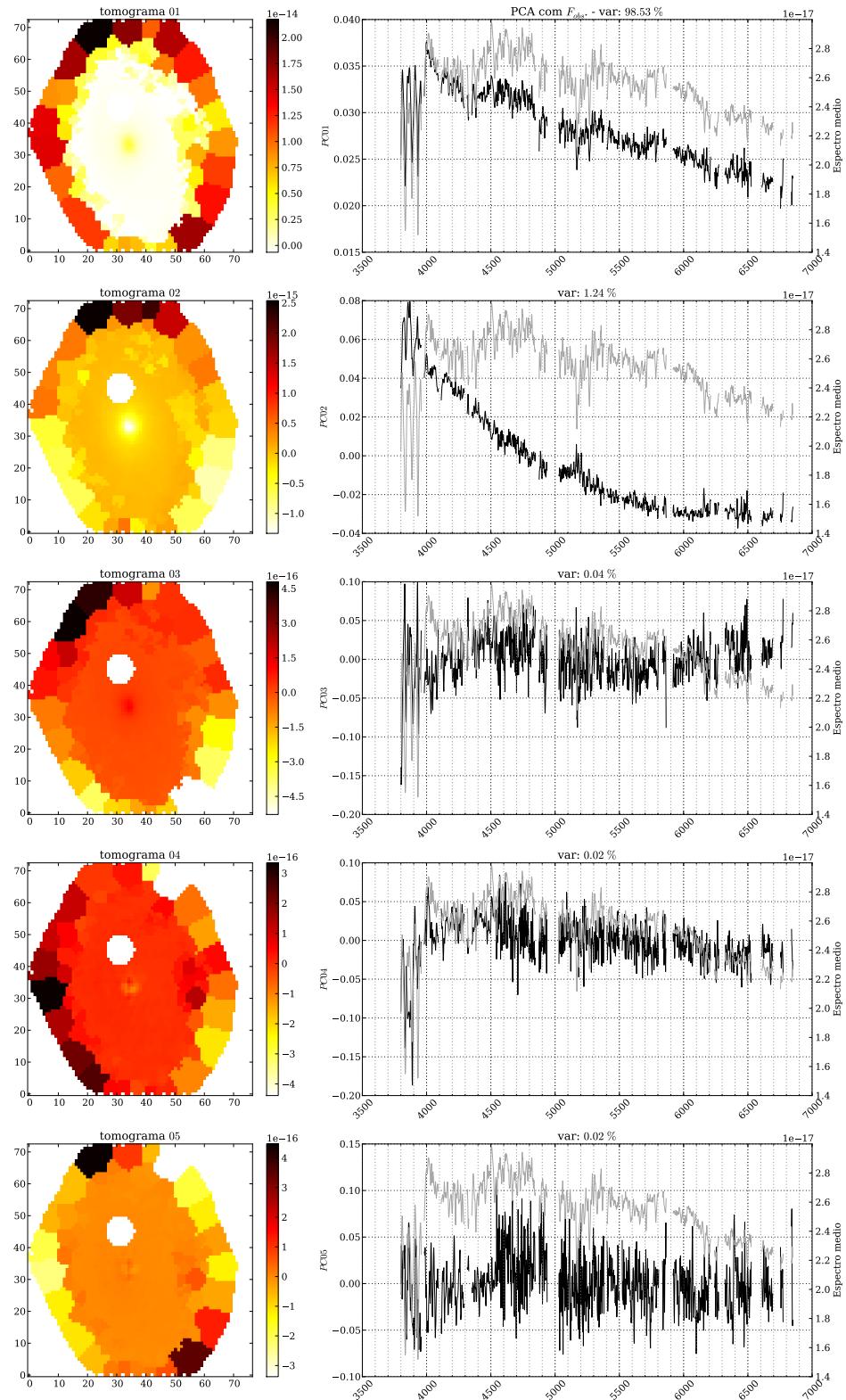


Figura 4.3: Cinco primeiras PCs (e seus respectivos tomogramas) resultantes da Tomografia PCA aplicado aos espectros sem normalização (F_{obs}) da galáxia NGC 2916.

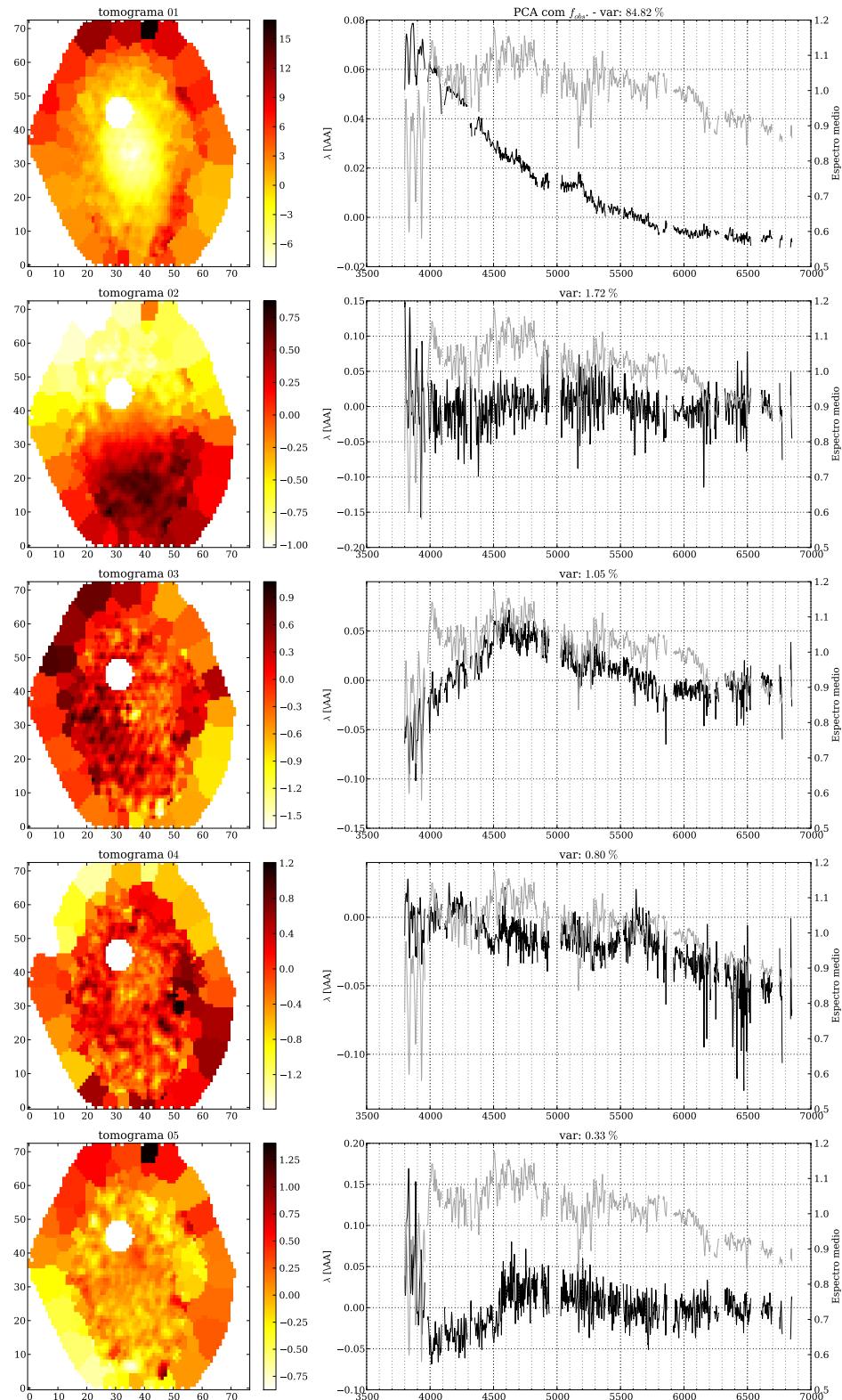


Figura 4.4: Cinco primeiras PCs (e seus respectivos tomogramas) resultantes da Tomografia PCA aplicado aos espectros observados normalizados (f_{obs}) da galáxia NGC 2916.

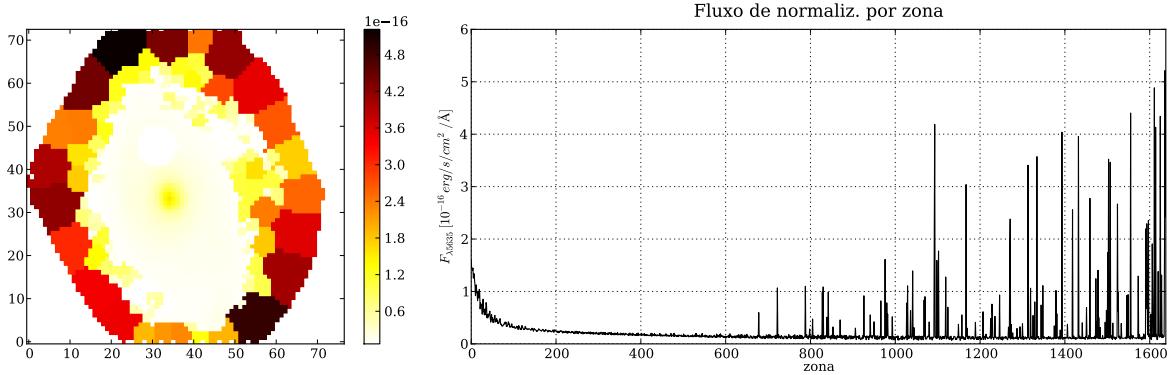


Figura 4.5: Fluxo usado para a normalização de cada espectro, mostrado tanto na forma de imagem (à esquerda) como em função do número da zona (direita).

4.2 PCA dos espectros sintéticos

Na seção anterior realizamos a 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 submetidas à técnica de PCA esperamos que as informações mais relevantes se condensem em menos PCs. Comparando os dois *scree tests* na Figura 4.6 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 (f_{syn}) da galáxia NGC 2916 aparecem na Figura 4.7. Comparando com seu correspondente observational na Figura 4.4, 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 populações estelares, não possuem nenhuma assinatura instrumental. Por esse fato acabamos descobrindo uma assinatura pre-

¹Efeitos de extinção/avermelhamento e cinemática também são incluídos no processo de síntese (?)

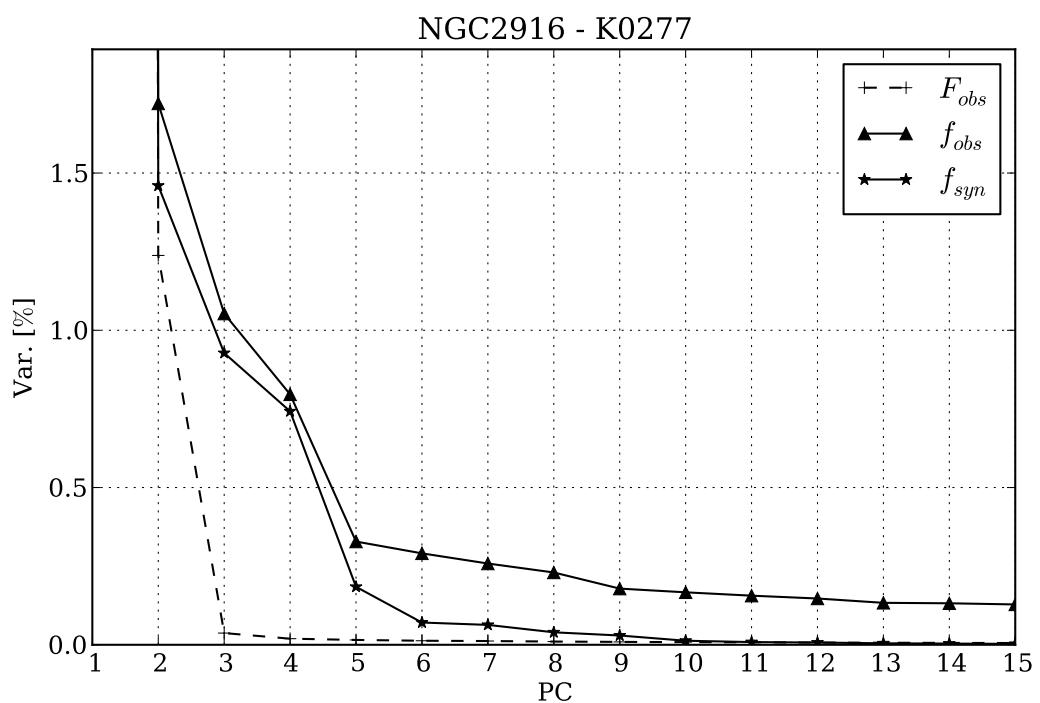


Figura 4.6: Scree test para 3 PCAs da galáxia NGC 2916 (CALIFA 277). Com marcações de triângulos vemos as PCs resultantes da PCA com os espectros observados normalizados (f_{obs}). As variâncias das PCs marcadas com estrela representam a PCA com os espectros sintéticos normalizados (f_{syn}). Para comparação plotamos as PCs do caso sem normalização (F_{obs}) usando linha pontilhada.

sente 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 4.4 mostra um degrau entre 4500 e 4600 \AA o 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 a 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.

²Componentes geradas pela 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.

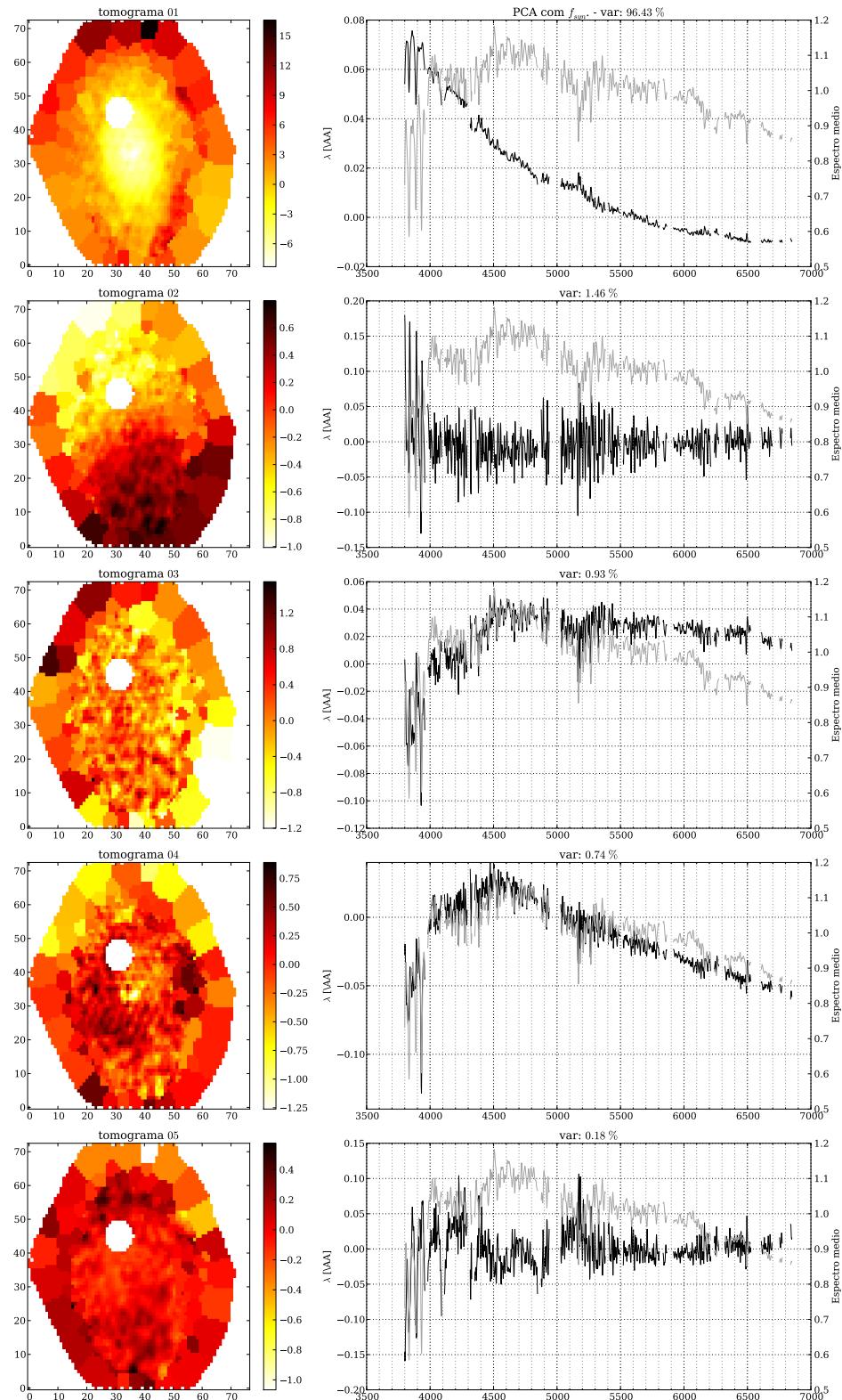


Figura 4.7: Cinco primeiras PCs (e seus respectivos tomogramas) resultantes da Tomografia PCA aplicado aos espectros sintéticos normalizados (f_{syn}) da galáxia NGC 2916.

4.3 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 pela PCA. Nas Figuras 4.8 e 4.9 vemos as correlações entre algumas propriedades físicas ($\langle \log t \rangle_L$, $\log \langle Z \rangle_L$, A_V , v_\star , σ_\star) 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_\star 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_\star . Por fim, a PC6 no caso observado parece ter uma mistura de idade, metalicidade, extinção e com a dispersão 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 Spearman 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 4.10 e 4.11 que mostram as PCs para o caso observado e sintético, coloridos pelo A_V .

A Figura 4.12 mostra também o efeito que a falta de normalização faz com o resultado da PCA. Esse fator de amplitude que não é removido pela normalização se concentra principal-

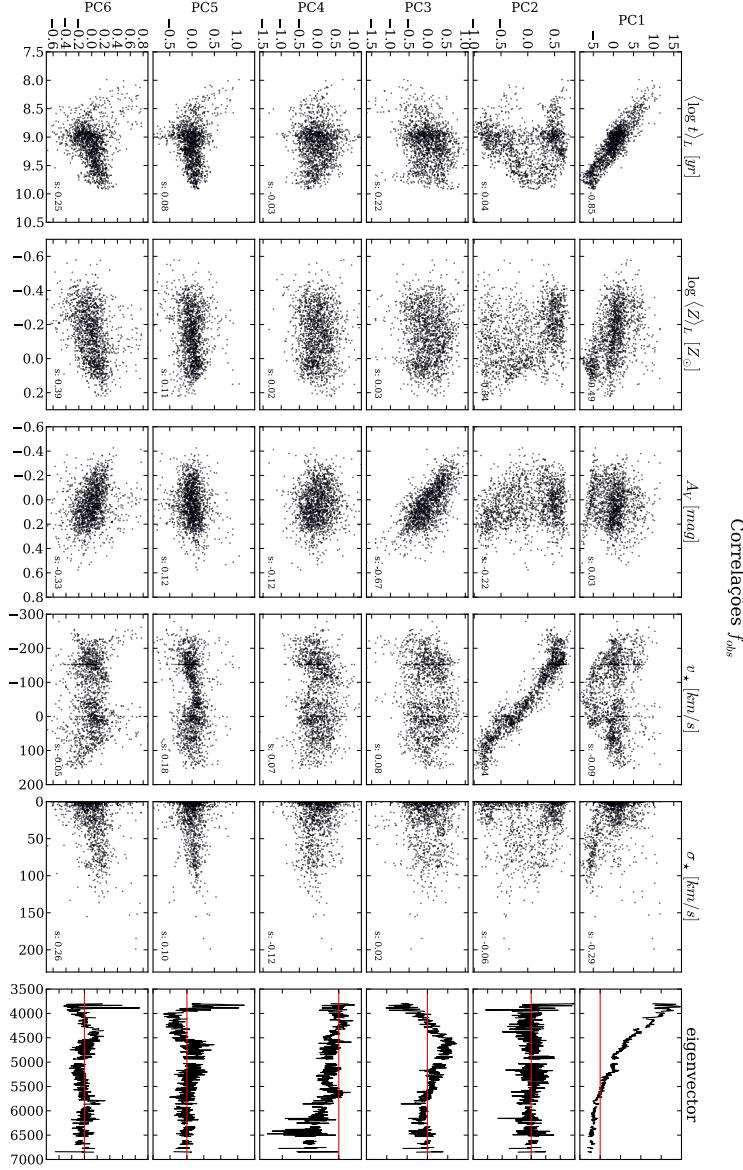


Figura 4.8: Correlações entre os pesos por zona das seis primeiras PCs da PCA feito para o cubo com os dados observados normalizados (f_{obs}) e cinco parâmetros físicos. Pela ordem de colunas da esquerda para direita temos $\langle \log t \rangle_L$, $\log \langle Z \rangle_L$, A_V , v_\star , σ_\star . 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.

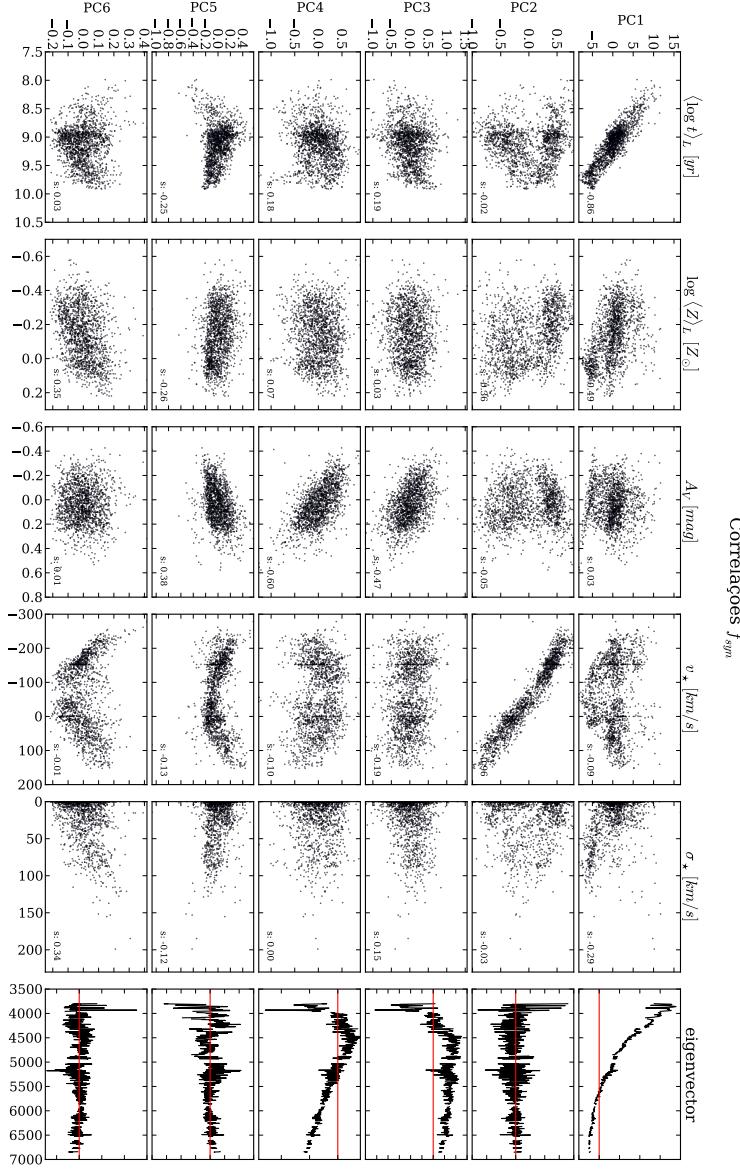


Figura 4.9: Correlações entre os pesos por zona das seis primeiras PCs da PCA feito para o cubo com os dados sintéticos normalizados (f_{syn}) e cinco parâmetros físicos. Pela ordem de colunas da esquerda para direita temos $\langle \log t \rangle_L$, $\log \langle Z \rangle_L$, A_V , v_\star , σ_\star . 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.

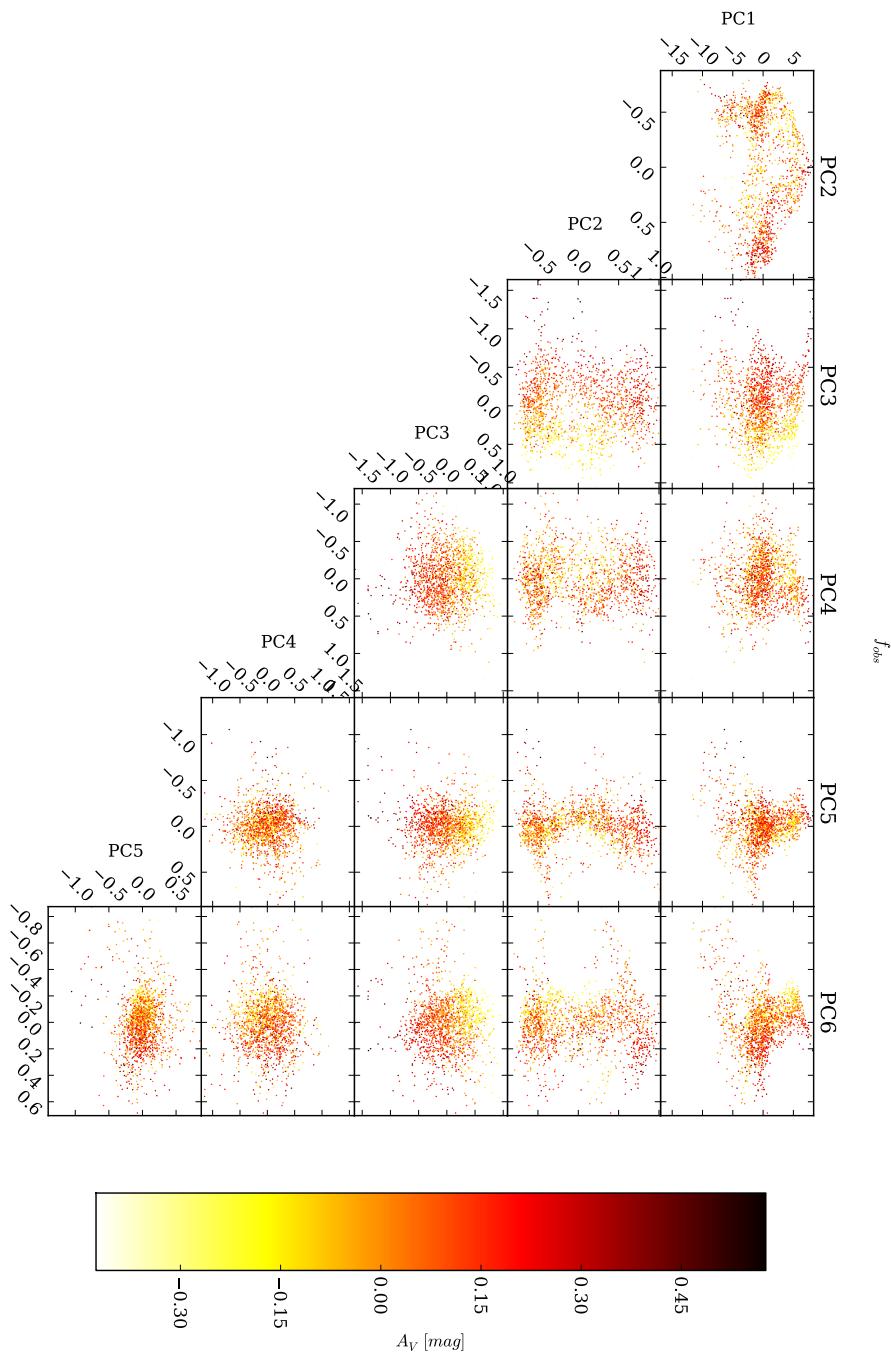


Figura 4.10: Dados das zonas no espaço das PCs (f_{obs}) coloridos pela extinção (A_V).

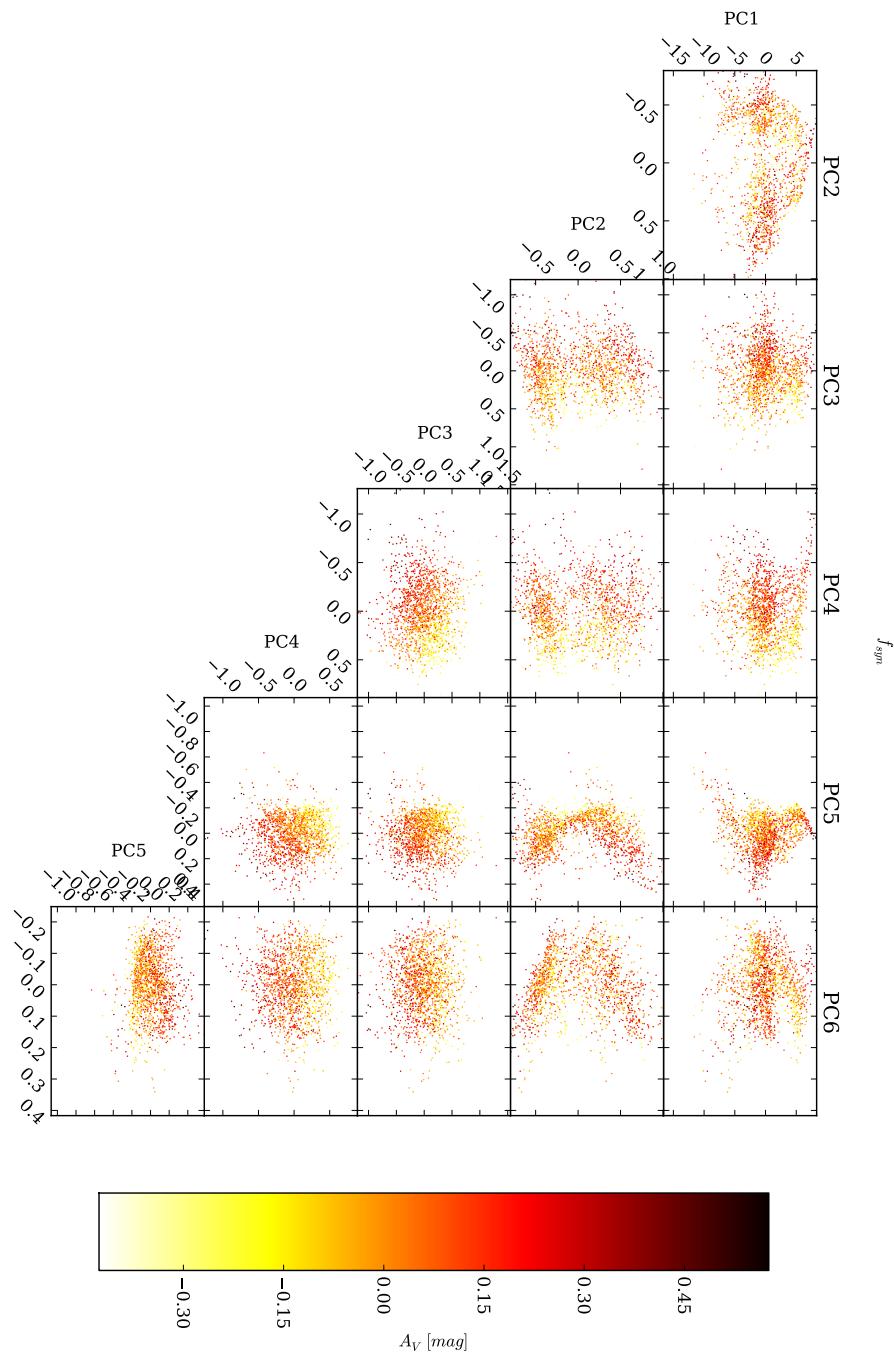


Figura 4.11: Dados das zonas no espaço das PCs (f_{syn}) coloridos pela extinção (A_V).

mente 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.

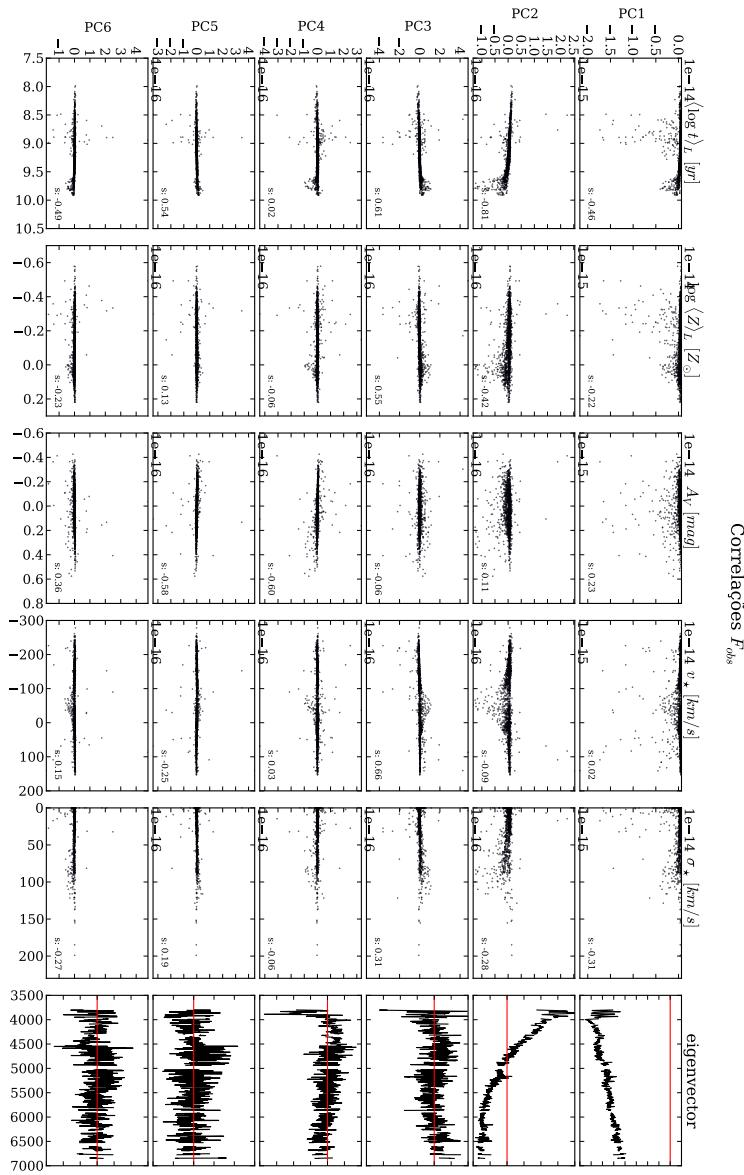


Figura 4.12: Correlações entre os pesos por zona das seis primeiras PCs da PCA feito para o cubo com os dados observados (F_{obs}) e cinco parâmetros físicos. Pela ordem de colunas da esquerda para direita temos $\langle \log t \rangle_L$, $\log \langle Z \rangle_L$, A_V , v_\star , σ_\star . Na última coluna temos o autoespectro para ajudar na visualização. A linha em vermelho no gráfico do autoespectro serve para identificar o zero.

Capítulo 5

Resultados: Tomografia PCA para 8 galáxias do CALIFA

No capítulo anterior apresentamos diferentes maneiras de se aplicar a PCA a cubos de dados, bem como uma técnica de “engenharia reversa”, na qual buscamos significados físicos das PCs através de correlações com propriedades das populações estelares obtidas pela síntese com o STARLIGHT. Este conjunto de técnicas nos dá munição para a primeira exploração científica usando Tomografia PCA em galáxias do CALIFA.

Nesse capítulo apresentamos resultados para uma amostra de 8 galáxias, escolhidas de forma a varrer vários tipos morfológicos. Essa é a primeira vez que esta técnica é aplicada a galáxias inteiras.

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 essas já se encontram no capítulo anterior. Todas as galáxias aqui estudadas estão listadas na Tabela 5.1 na ordem que aparecerão neste capítulo.

Para cada galáxia, apresentaremos inicialmente sua imagem SDSS acompanhada de produtos do QBICK e PyCASSO: mapas do fluxo em 5635 Å, a idade média ponderada pelo fluxo nesse comprimento de onda ($\langle \log t \rangle_L$), o logarítmico da metalicidade média ponderada pelo fluxo ($\log \langle Z \rangle_L$), a extinção estelar (A_V), o campo de velocidades na linha de visada (v_\star) e

a dispersão de velocidades (σ_\star). Em seguida, apresentamos o *scree test* para três tipos de PCA. Na sequência, apresentamos tomogramas e autoespectros para a PCA com o espectro observado normalizado e com o espectro sintético normalizado (como nas Figuras 4.4 e 4.7 do Capítulo 4). Finalmente, apresentamos correlações entre as PCs e propriedades derivadas pelo STARLIGHT (como nas Figuras 4.8 e 4.9 do Capítulo 4). Referência a PCA usando espectros observados será feita através da variável F_{obs} , usando os espectros observados normalizados, f_{obs} e usando os espectros sintéticos normalizados, f_{syn} . Algumas vezes nos referenciamos como caso observado e caso sintético para as análises PCA com os espectros normalizados, observados ou sintéticos.

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*, a distância correspondente a um spaxel do CALIFA (1 arcsec²) e número de zonas.

Nome da galáxia	CALIFA ID	<i>Hubble Type</i>	$\log M_\star [M_\odot]$	redshift	1 spaxel ~ [pc]	N_z
NGC 2916	K0277	Sbc	10.83	0.01244	248	1638
NGC 0001	K0008	Sbc	11.00	0.01515	302	1132
NGC 0776	K0073	SBb	11.19	0.01640	327	1733
NGC 4210	K0518	SBb	10.49	0.00906	181	1938
NGC 1167	K0119	S0	11.47	0.01645	328	1879
NGC 6515	K0864	E3	11.42	0.02285	455	887
NGC 2623	K0213	Scd	10.74	0.01847	368	561
ARP 220	K0802	Sd	11.15	0.01814	361	1157

5.2 Galáxias espirais

5.2.1 NGC 0001 - CALIFA 8

No capítulo anterior escolhemos como exemplo a galáxia NGC 2916 pelo fato que é uma das galáxias mais estudadas pelo nosso grupo até agora (??). 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 4.2, onde o resultado do PCA para f_{syn} 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} , através da nossa análise dos tomogramas as componentes daí pra frente não parecem ter mais informações “decifráveis”.

Através da análise das PCs para o caso observado dessa galáxia (Figura 5.3), notamos que a parte inicial (os primeiros ~ 40 Å) e a parte final (os últimos ~ 10 Å) dos autoespectros chamam atenção e, através de uma revisão dos espectros do cubo, vimos que elas, na verdade, evidenciam a existência problemas com o fluxo nesses comprimentos de onda em alguns espectros. Mesmo assim, seguimos com nossa análise. Como a NGC 2916, a primeira PC, e seu respectivo tomograma (Figuras 5.3 e 5.4), parece ser relacionada a uma população muito azul, tanto para a 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 5.1, 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 espiral barrada NGC 0776 é apresentada na Figura 5.7. A análise com o STARLIGHT revela populações mais jovens nos braços espirais, uma população velha na região da barra e levemente mais jovem no núcleo. A imagem de A_V mostra que a poeira se concentra no núcleo e nos braços. O campo de velocidade mostra pequenos valores de v_* , o que não é supreendente dado que a galáxias está praticamente *face-on*.

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

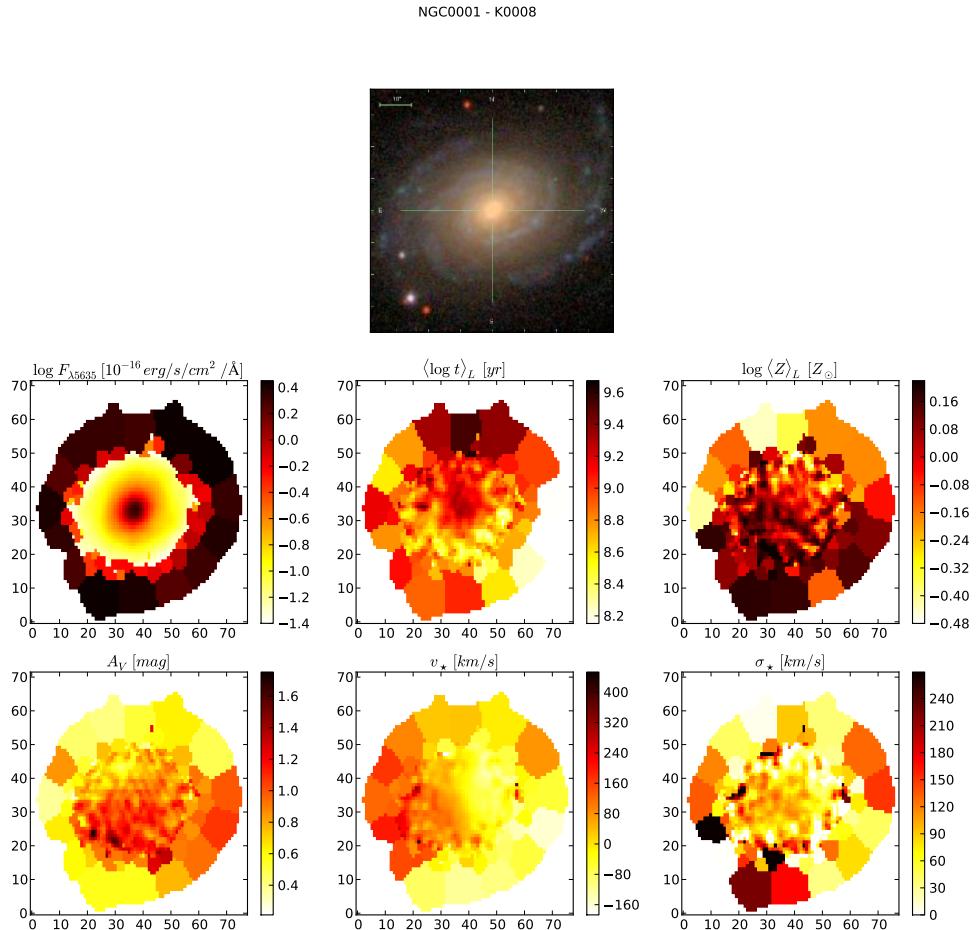


Figura 5.1: Propriedades físicas da galáxia NGC 0001. Na primeira linha a imagem do SDSS da galáxia. Na primeira imagem da segunda linha temos o valor do fluxo observado em 5635 Å, usado para a normalização dos espectros de cada zona. Da segunda imagem da segunda linha em diante temos as propriedades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

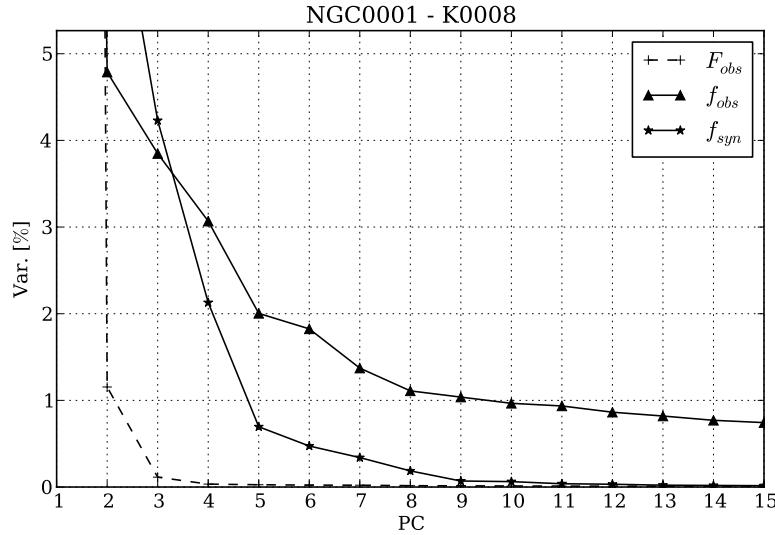


Figura 5.2: Scree test para 3 análises PCA da galáxia NGC 0001 (CALIFA 8). Com marcações de triângulos vemos as PCs resultantes do PCA com os espectros observados normalizados (f_{obs}). As variâncias das PCs marcadas com estrela representam o PCA com os espectros sintéticos normalizados (f_{syn}). Para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

de variância seguindo a ordem primeiro F_{obs} , depois f_{syn} e por último f_{obs} . 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). De fato o tomograma da PC1 se assemelha bastante a imagem da idade média. 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.

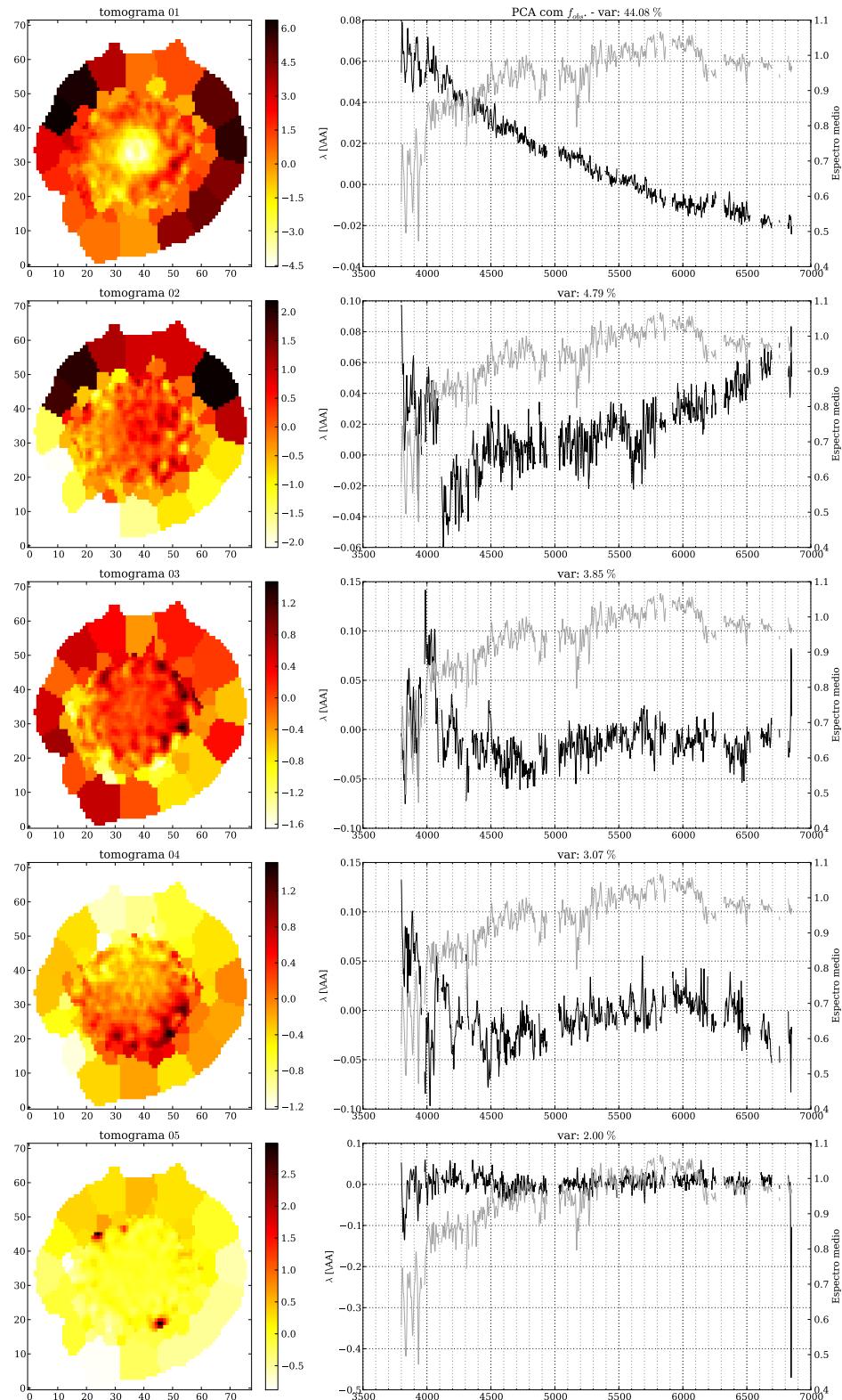


Figura 5.3: Cinco primeiras PCs (e seus respectivos tomogramas) resultantes da Tomografia PCA aplicado aos espectros observados normalizados (f_{obs}) da galáxia NGC 0001.

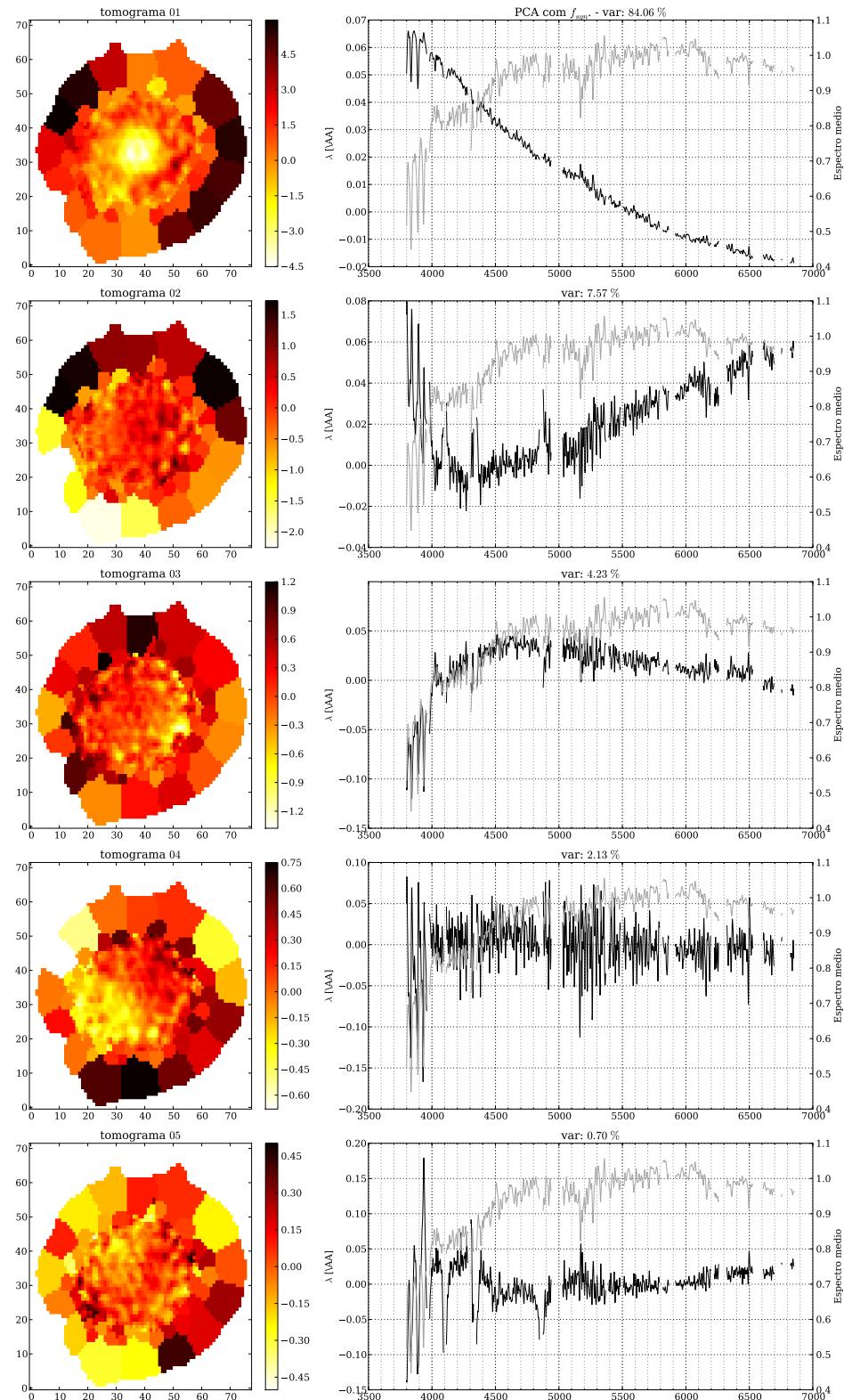


Figura 5.4: Cinco primeiras PCs (e seus respectivos tomogramas) resultantes da Tomografia PCA aplicado aos espectros sintéticos normalizados (f_{sym}) da galáxia NGC 0001.

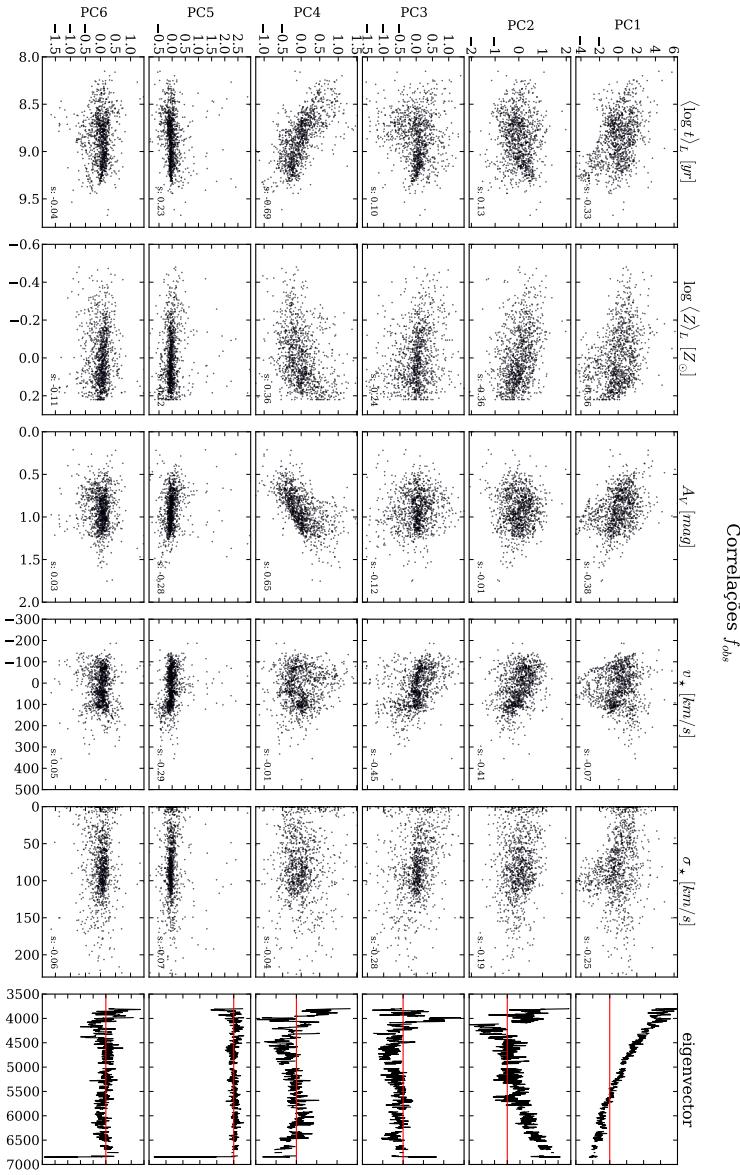


Figura 5.5: Correlações entre os pesos por zona das seis primeiras PCs do PCA feito para o cubo com os espectros observados normalizados (f_{obs}) e cinco parâmetros físicos. Pela ordem de colunas da esquerda para direita temos $\langle \log t \rangle_L$, $\log \langle Z \rangle_L$, A_V , v_* , σ_* . 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.

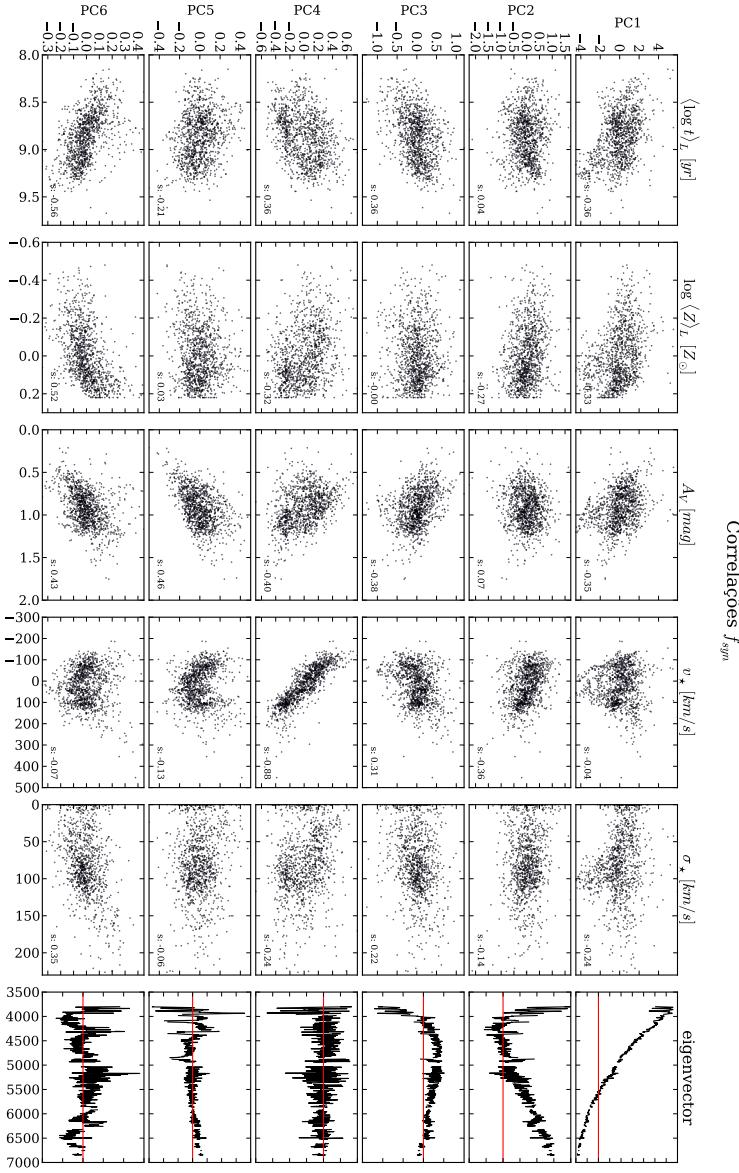


Figura 5.6: Correlações entre os pesos por zona das seis primeiras PCs do PCA feito para o cubo com os espectros sintéticos normalizados (f_{syn}) e cinco parâmetros físicos. Pela ordem de colunas da esquerda para direita temos $\langle \log t \rangle_L$, $\log \langle Z \rangle_L$, A_V , v_\star , σ_\star . 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.

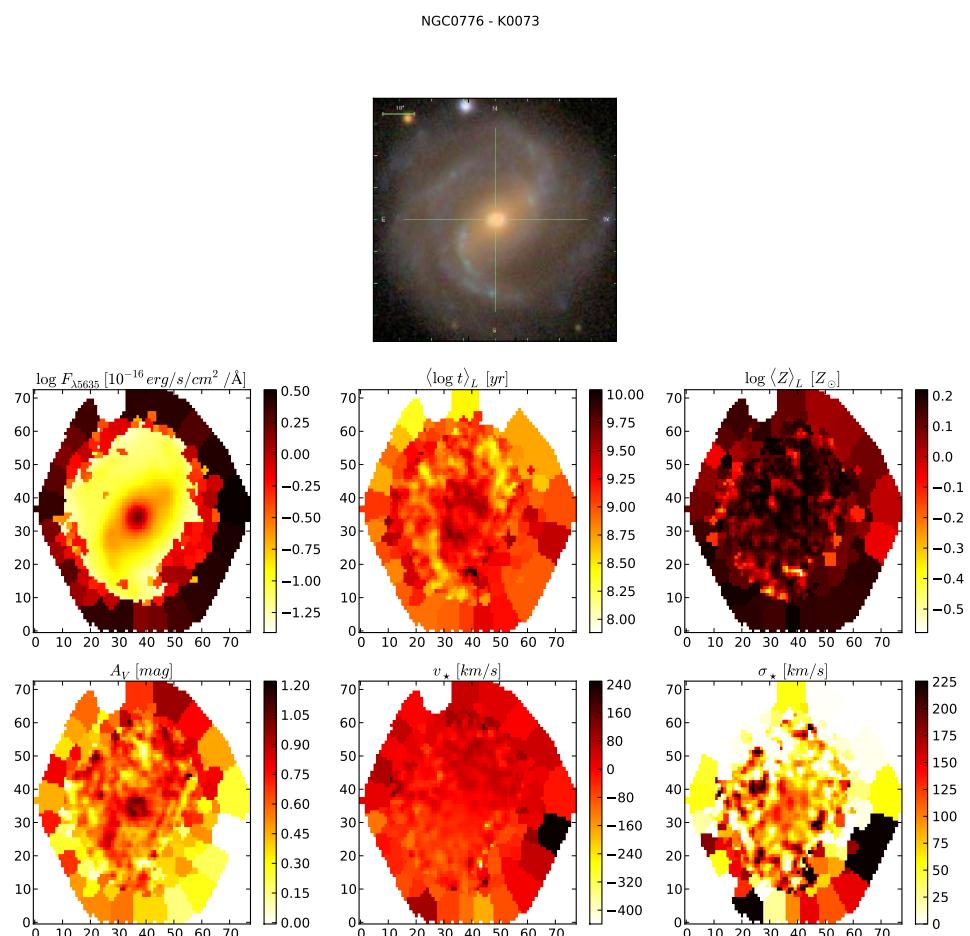


Figura 5.7: Igual a Figura 5.1 para a galáxia NGC 0776.



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

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 3 conjuntos aproximadamente circulares de pixels à direita do núcleo. Evidentemente existem problemas com esses pixels e esses problemas se propagaram 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 e os filtros de qualidade do QBICK. Ao invés de voltar atrás e corrigir esses defeitos, usaremos esse cubo como exemplo de como o PCA pode nos ajudar a identificar problemas nos dados.

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 para o caso observado vemos também um exemplo de “linha larga” ($\sim 6000 - 6200 \text{ \AA}$) que também foi causado por defeitos em zonas com espectros defeituosos em nesses 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).

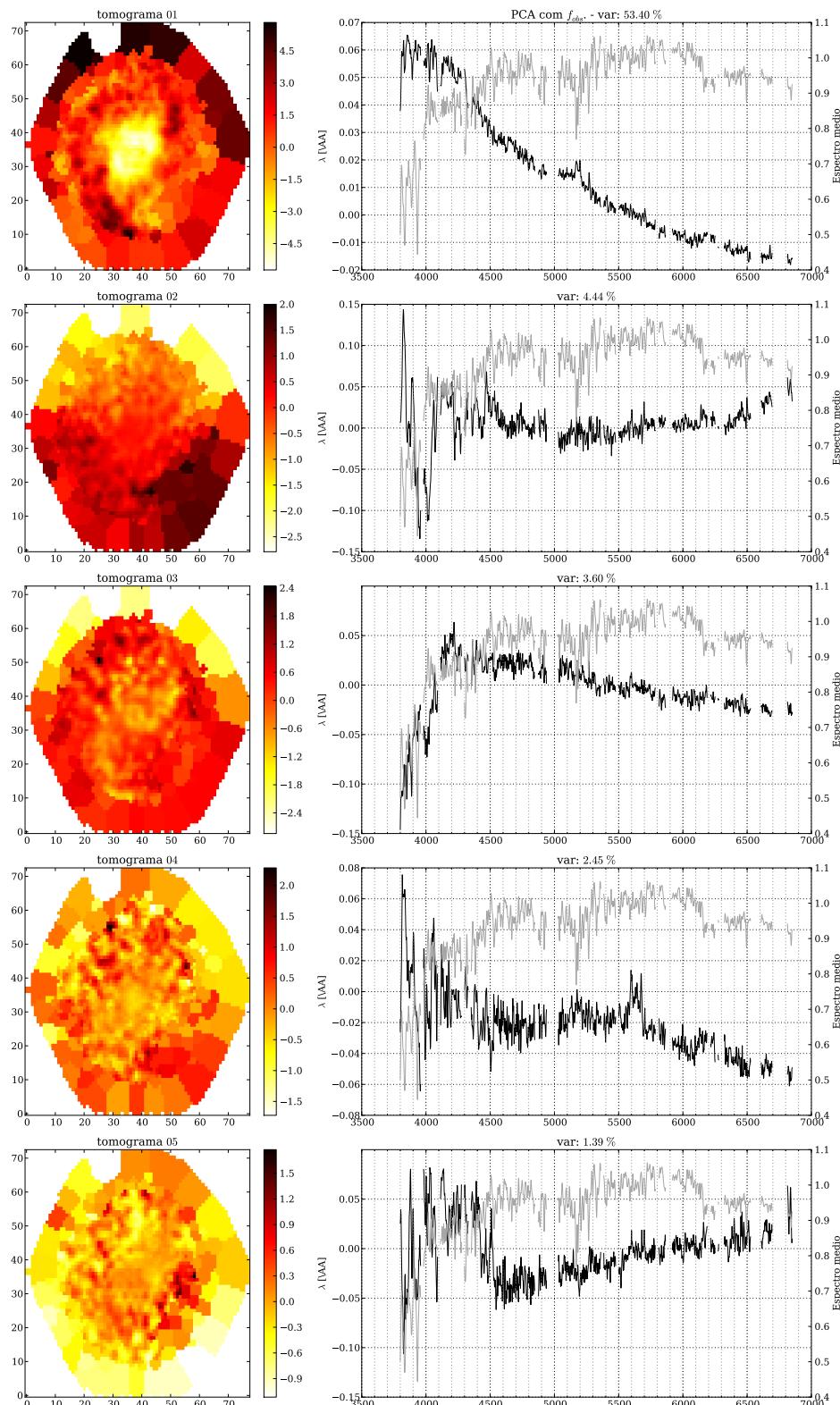


Figura 5.9: Igual a Figura 5.3 para a galáxia NGC 0776.

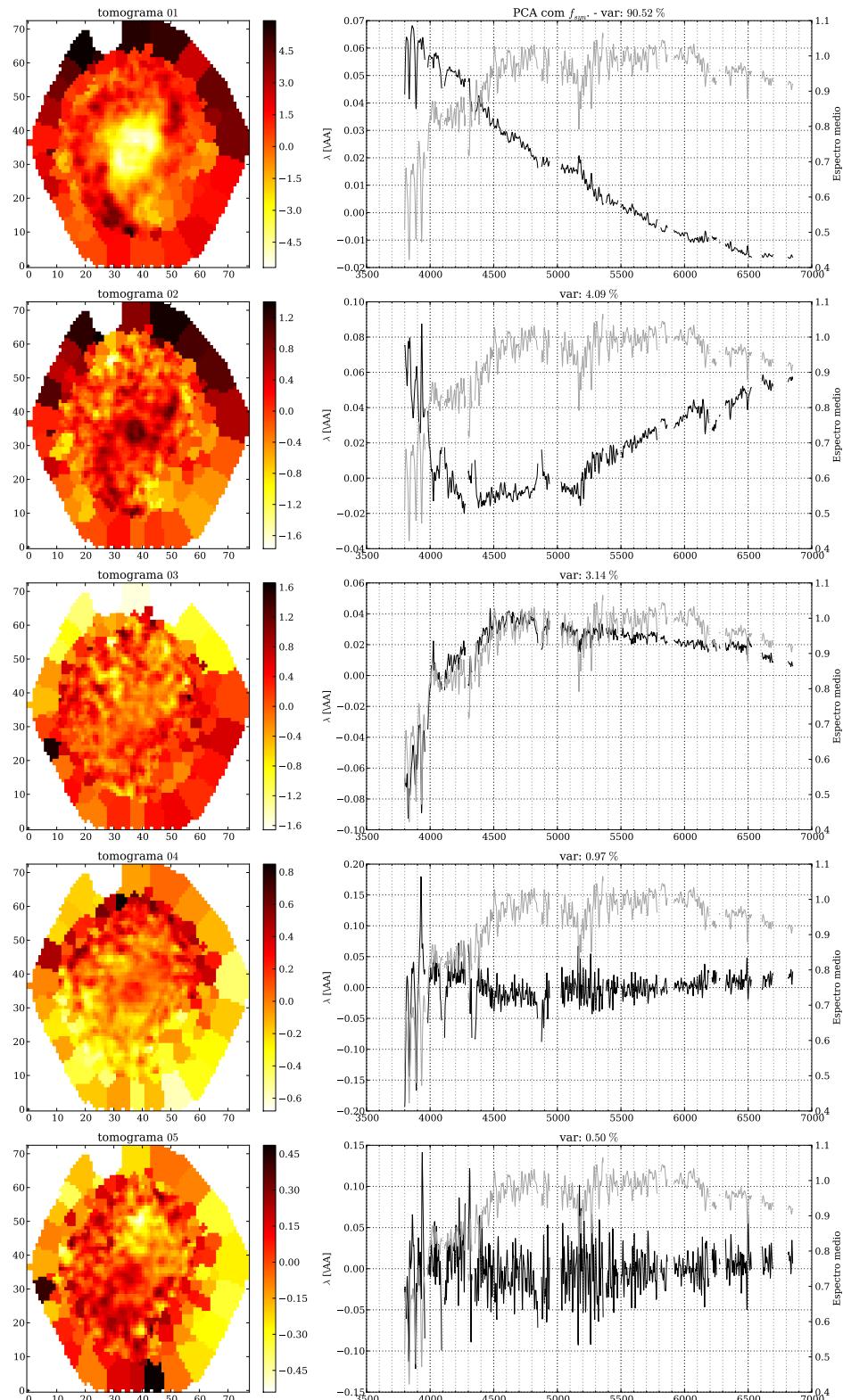


Figura 5.10: Igual a Figura 5.4 para a galáxia NGC 0776.

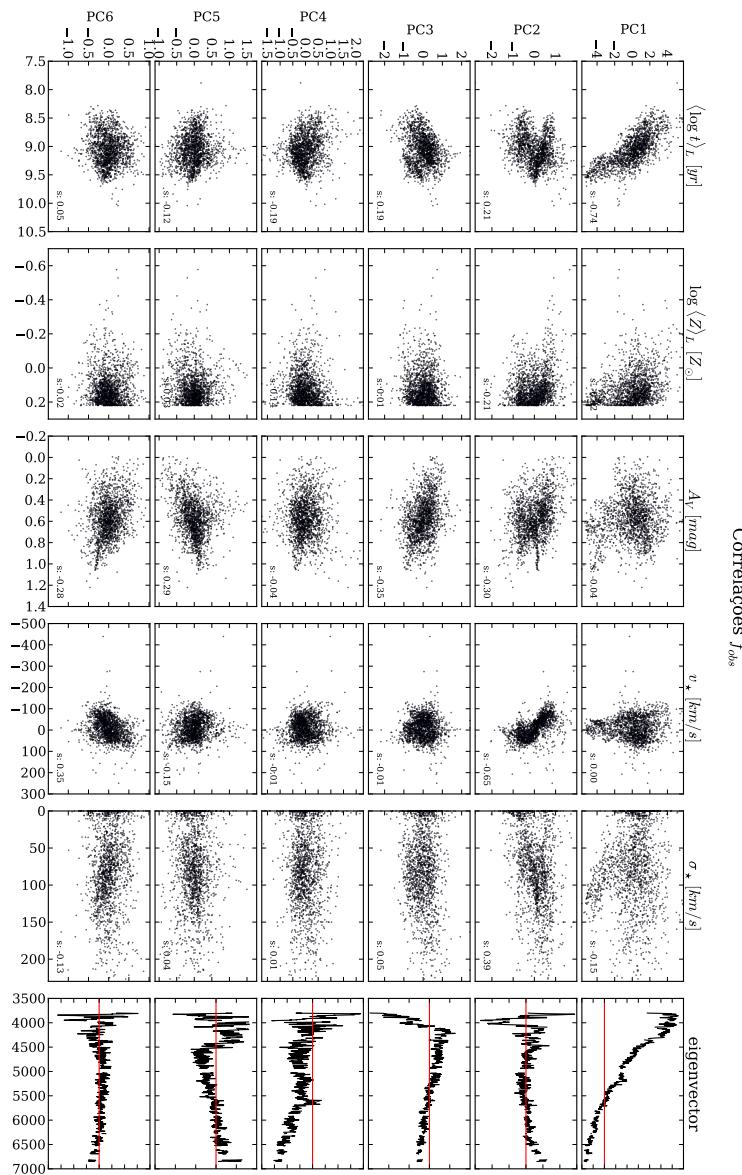


Figura 5.11: Igual a Figura 5.5 para a galáxia NGC 0776.

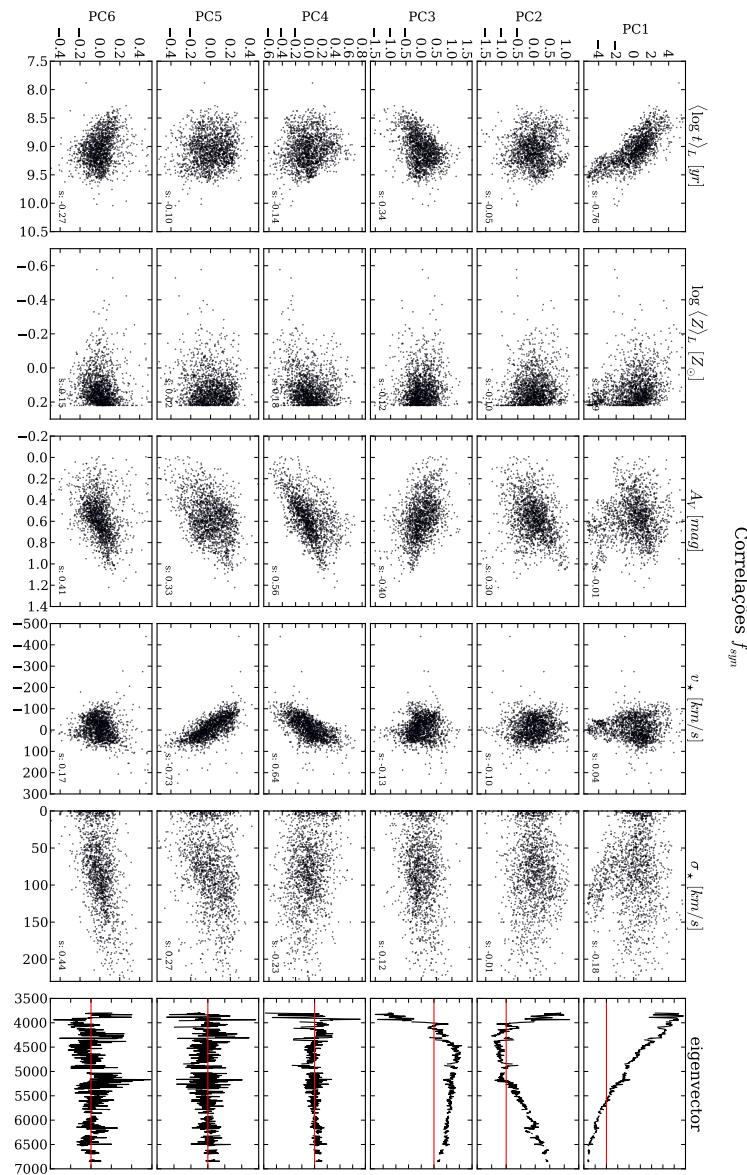


Figura 5.12: Igual a Figura 5.6 para a galáxia NGC 0776.

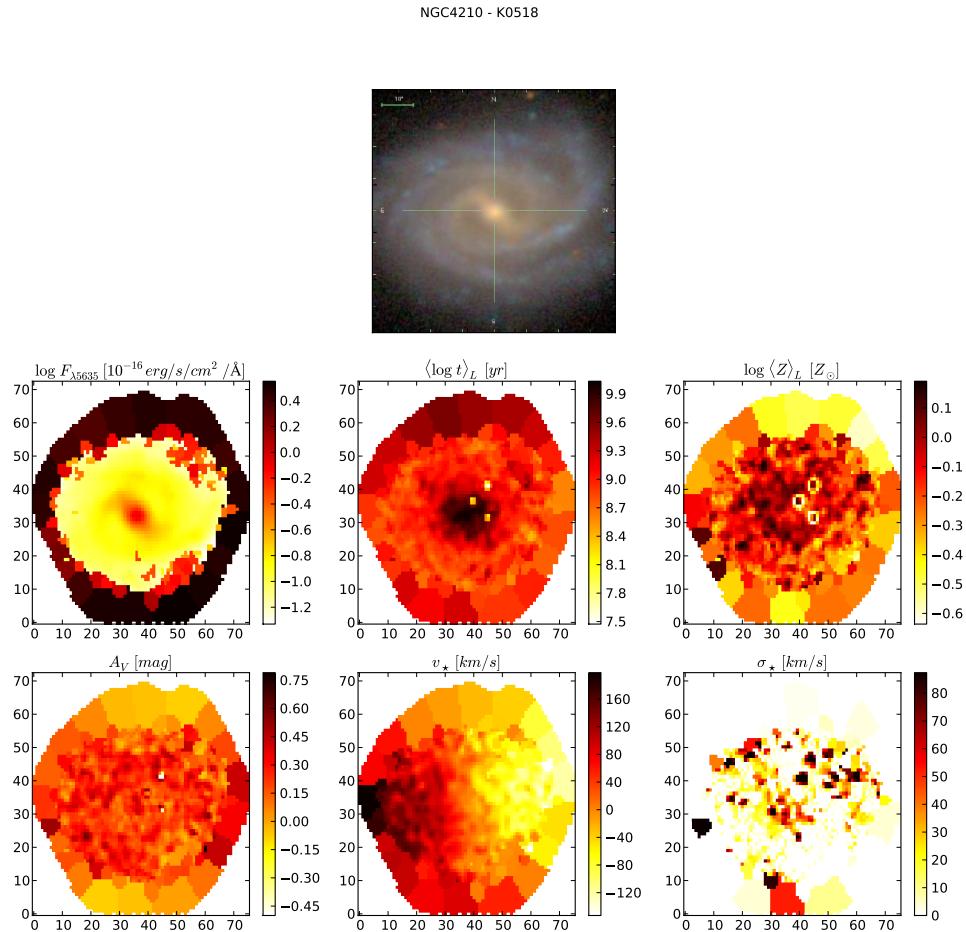


Figura 5.13: Igual a Figura 5.1 para a galáxia NGC 4210.

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 caso observado (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 5.14: Igual a Figura 5.2 para a galáxia NGC 4210.

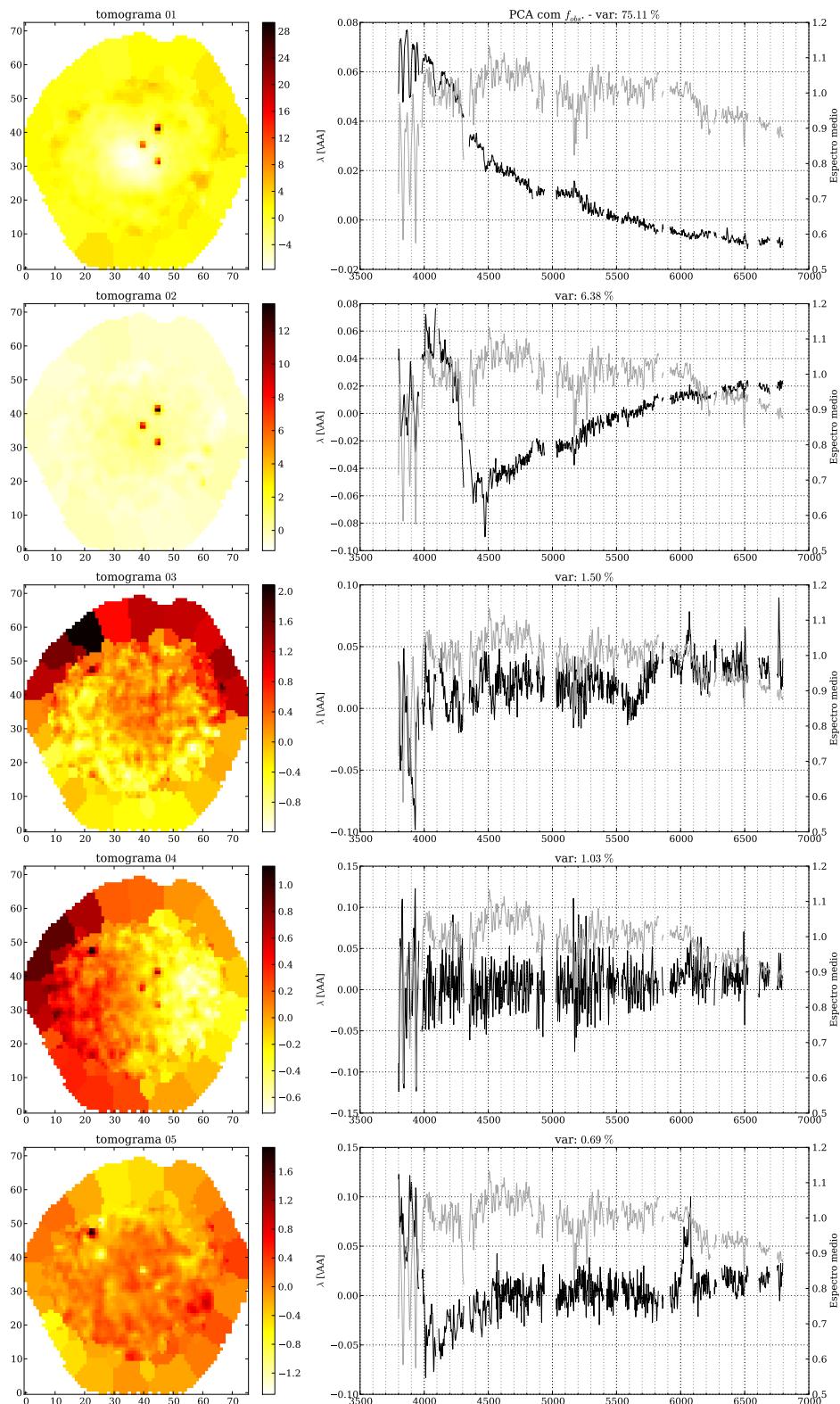


Figura 5.15: Igual a Figura 5.3 para a galáxia NGC 4210.

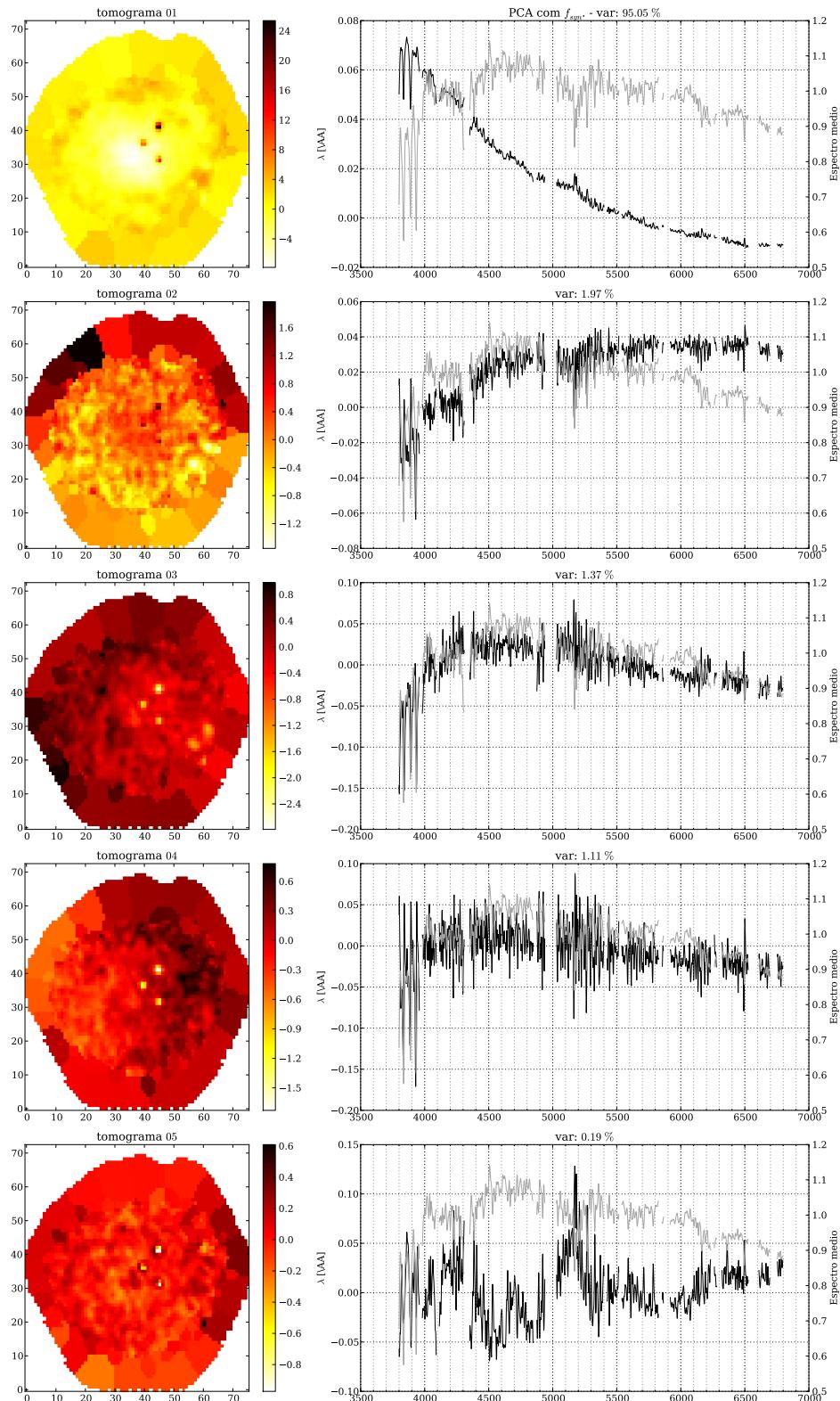


Figura 5.16: Igual a Figura 5.4 para a galáxia NGC 4210.

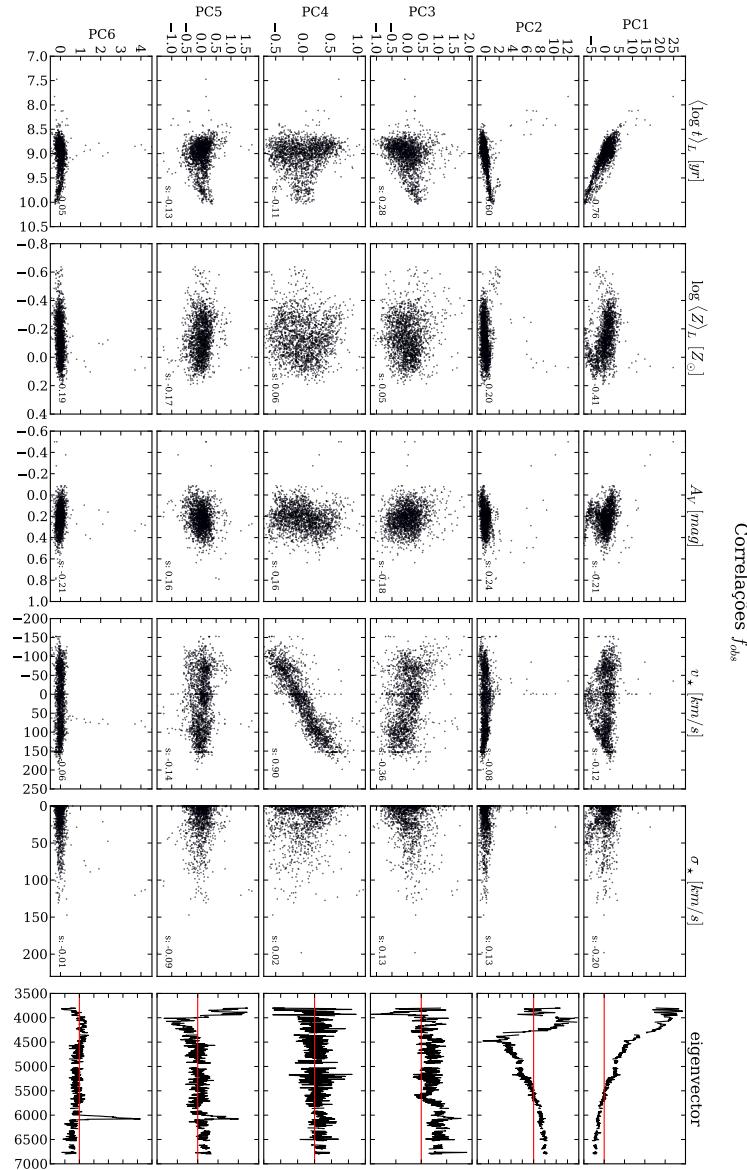


Figura 5.17: Igual a Figura 5.5 para a galáxia NGC 4210.

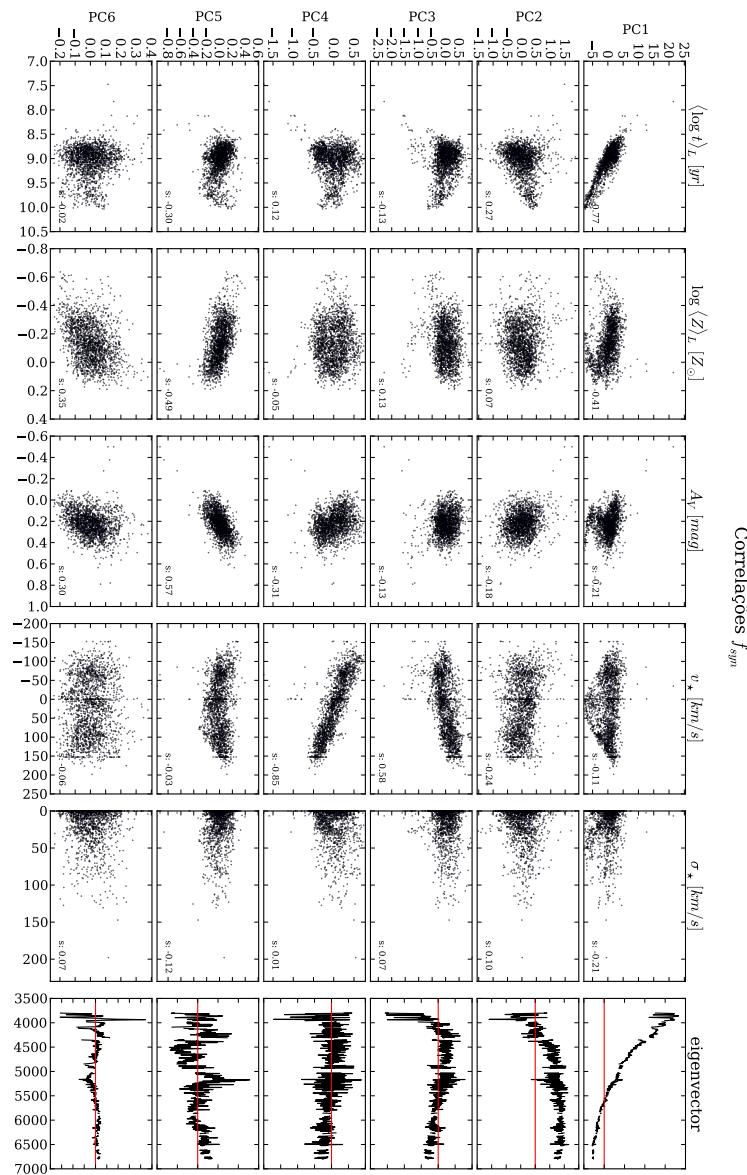


Figura 5.18: Igual a Figura 5.6 para a galáxia NGC 4210.

5.3 Galáxias early-type

Galáxias elípticas e S0 são sabidamente mais simples do que espirais, tanto em termos de estruturas morfológicas como em suas populações estelares, mais velhas do que aquelas encontradas nos discos (e as vezes nas regiões centrais) de espirais. Elas também contêm menos poeira, sendo essa também distribuída de forma mais regular. Esta maior homogeneidade das propriedades dessas galáxias sugere uma menor variância espectral. Vejamos o que a tomografia PCA revela em duas 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 dispersão 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 ela exibe uma ótima correlação com A_V e com a 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 incrível índice de correlação de 0.98 para o caso sintético. Vemos que a dispersão de velocidades (σ_\star) aparece correlacionando fracamente com quase todas as principais PCs, mas no caso sintético vemos uma correlação maior ($s = 0.53$). 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 elíptica é o número de zonas faltantes na análise. Existem muitos objetos no *FoV*

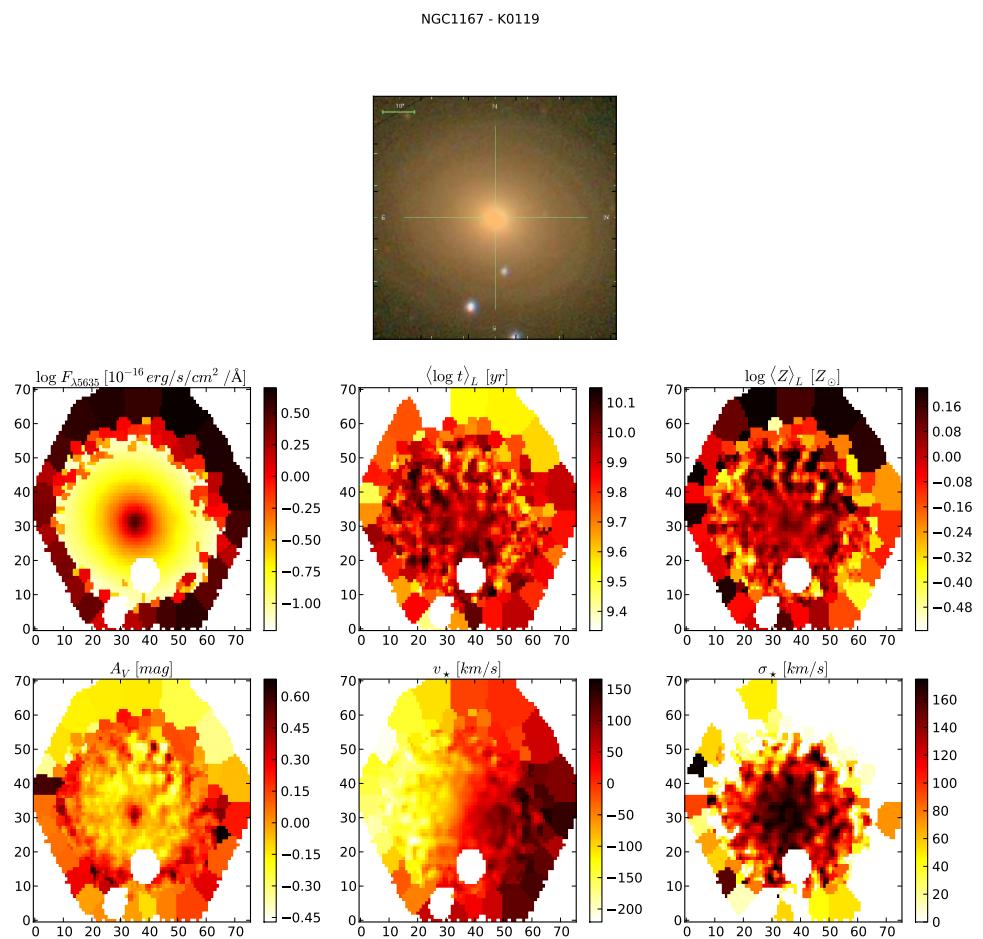


Figura 5.19: Igual a Figura 5.1 para a galáxia NGC 1167.



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

de observação dessa galáxia, portanto esses “buracos” são causados pelas máscaras espaciais criadas pelo QBICK conforme discutido na Seção 2.2. 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 dispersão de velocidades a correlação é um pouco mais alta ($s = 0.62$). A PC4 possui o mesmo padrão de correlações distribuídas 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.

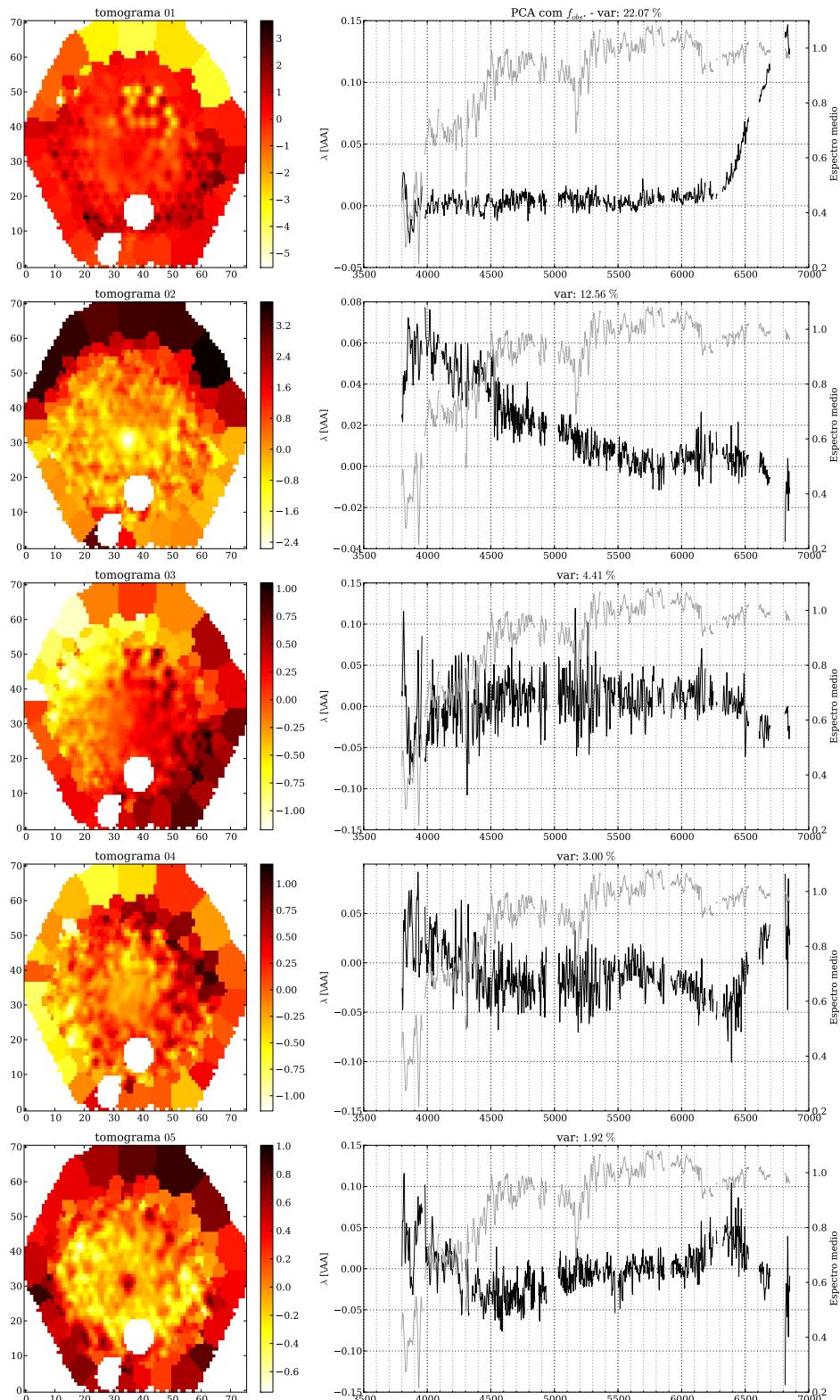


Figura 5.21: Igual a Figura 5.3 para a galáxia NGC 1167.

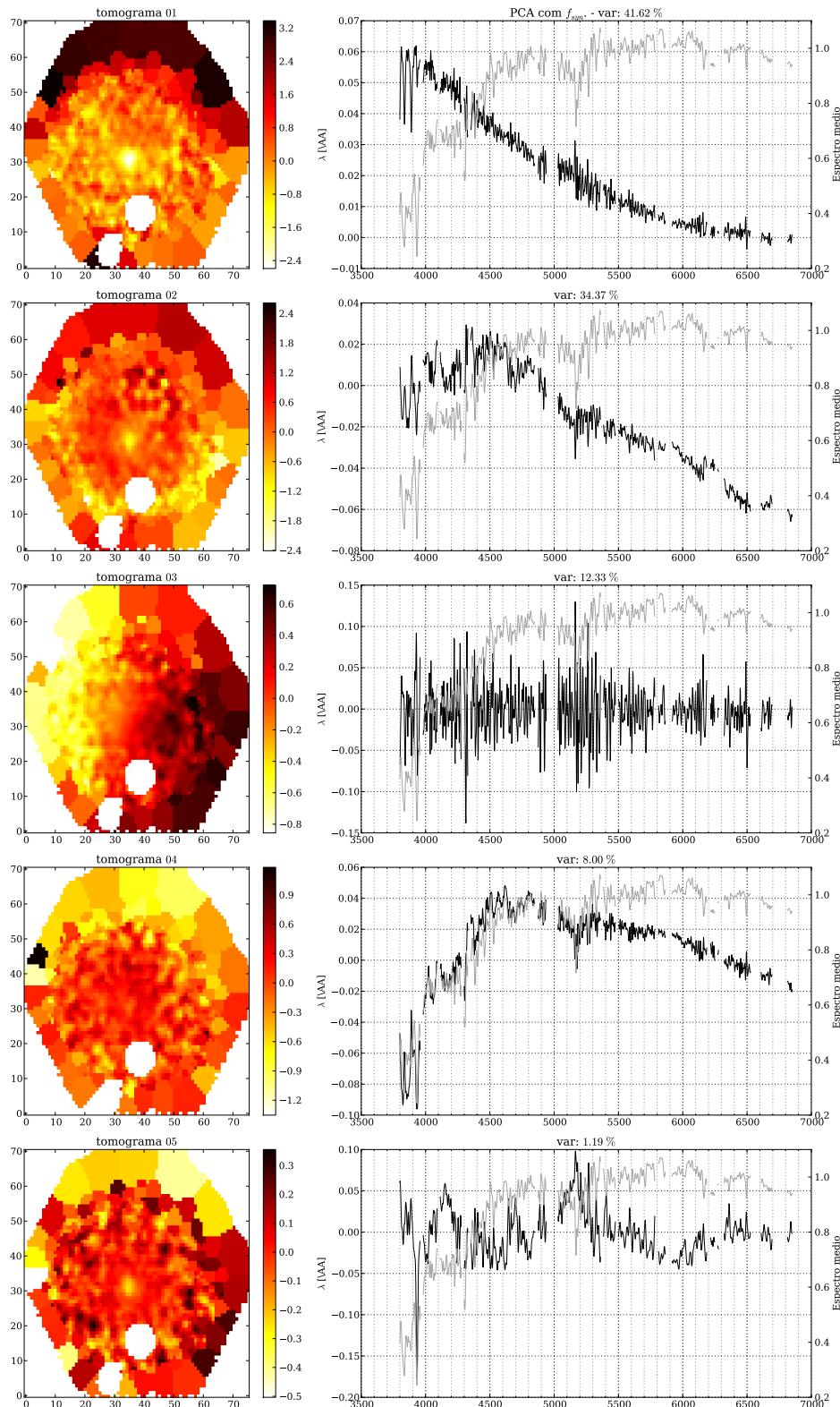


Figura 5.22: Igual a Figura 5.4 para a galáxia NGC 1167.

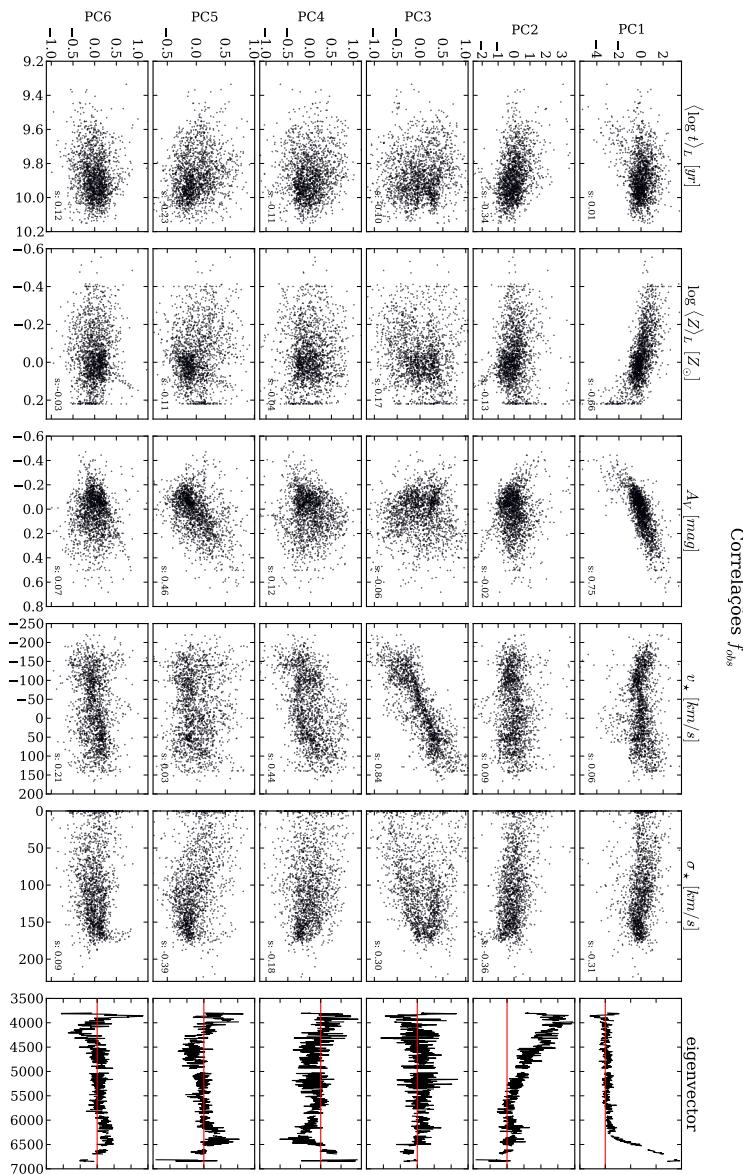


Figura 5.23: Igual a Figura 5.5 para a galáxia NGC 1167.

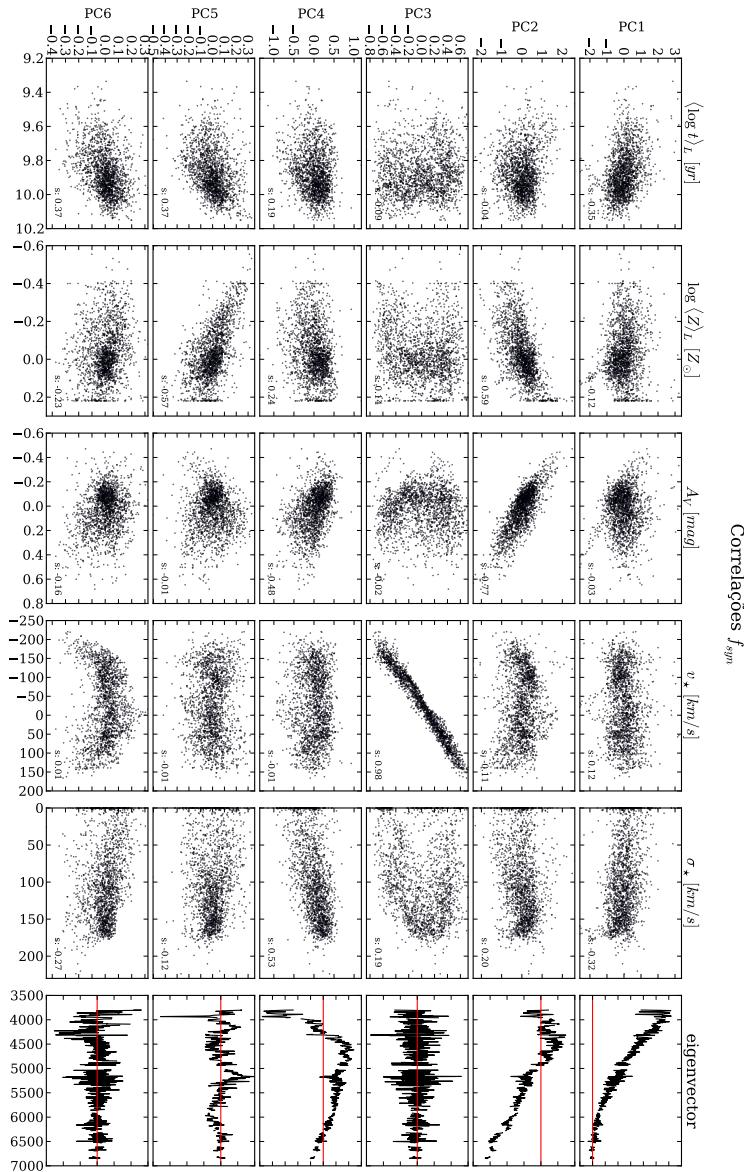


Figura 5.24: Igual a Figura 5.6 para a galáxia NGC 1167.

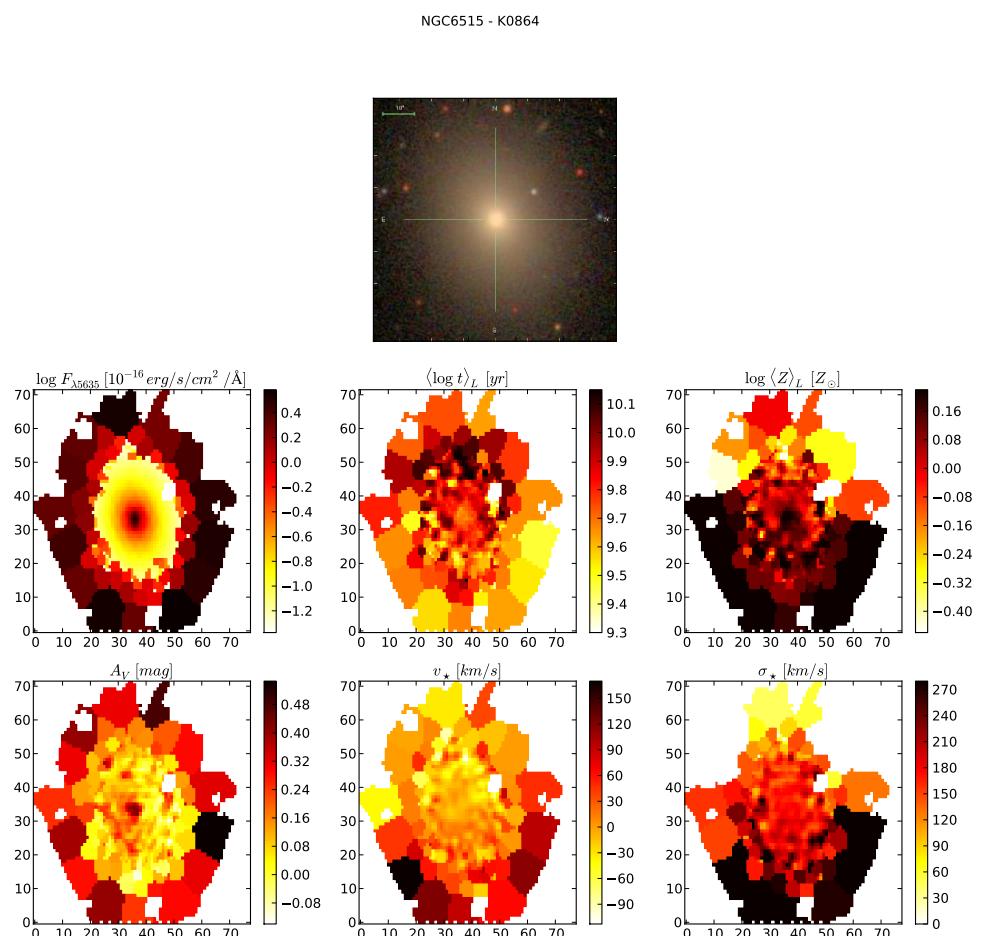


Figura 5.25: Igual a Figura 5.1 para a galáxia NGC 6515.



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

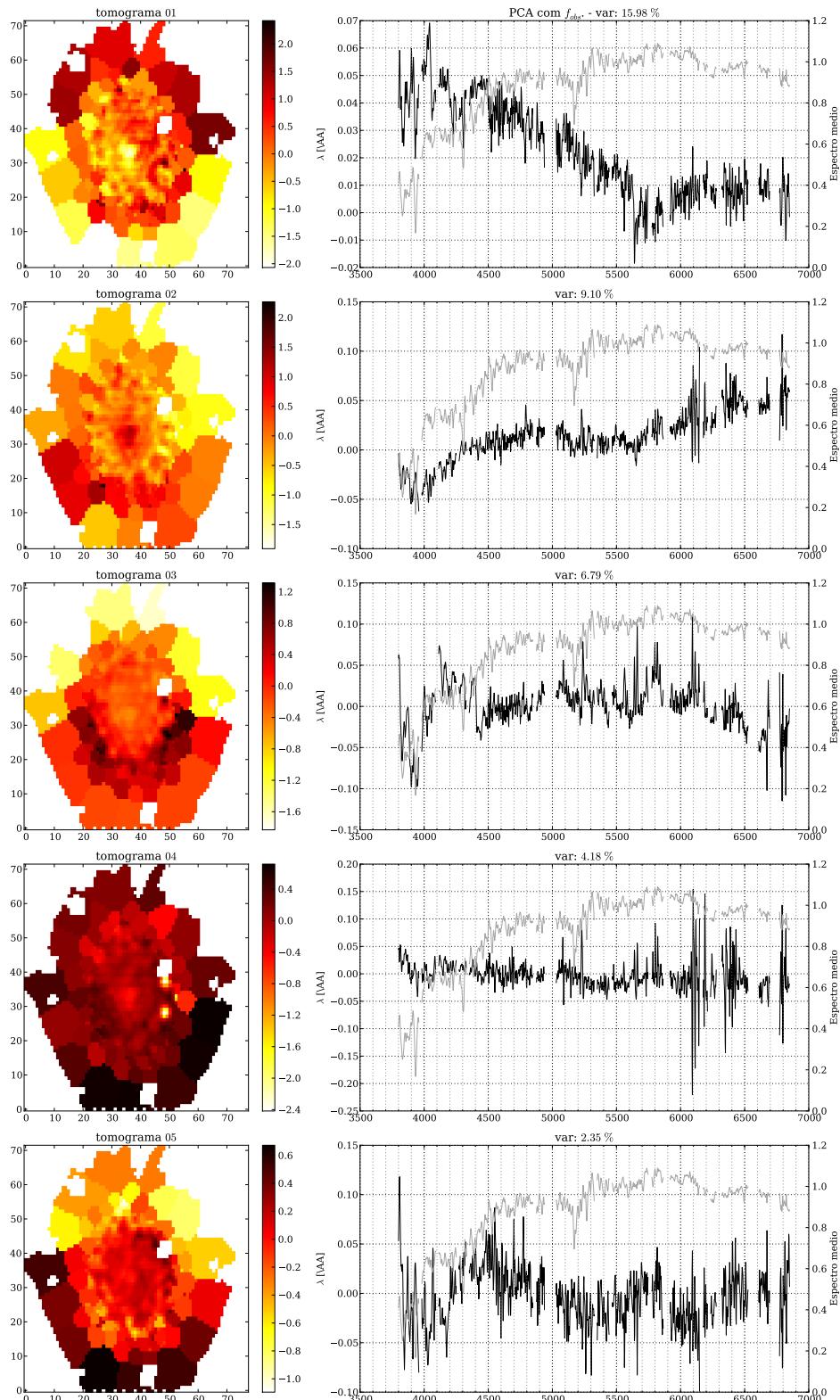


Figura 5.27: Igual a Figura 5.3 para a galáxia NGC 6515.

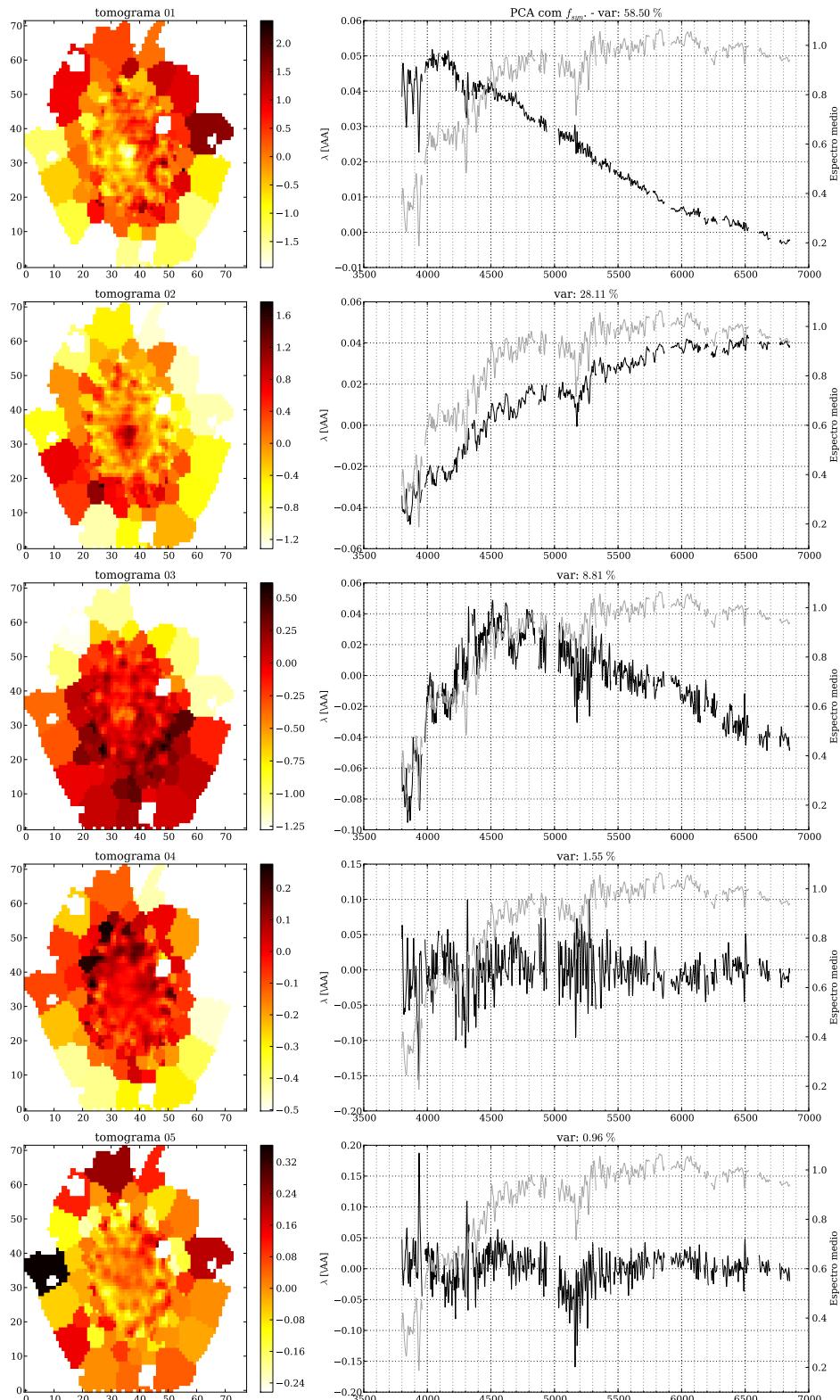


Figura 5.28: Igual a Figura 5.4 para a galáxia NGC 6515.

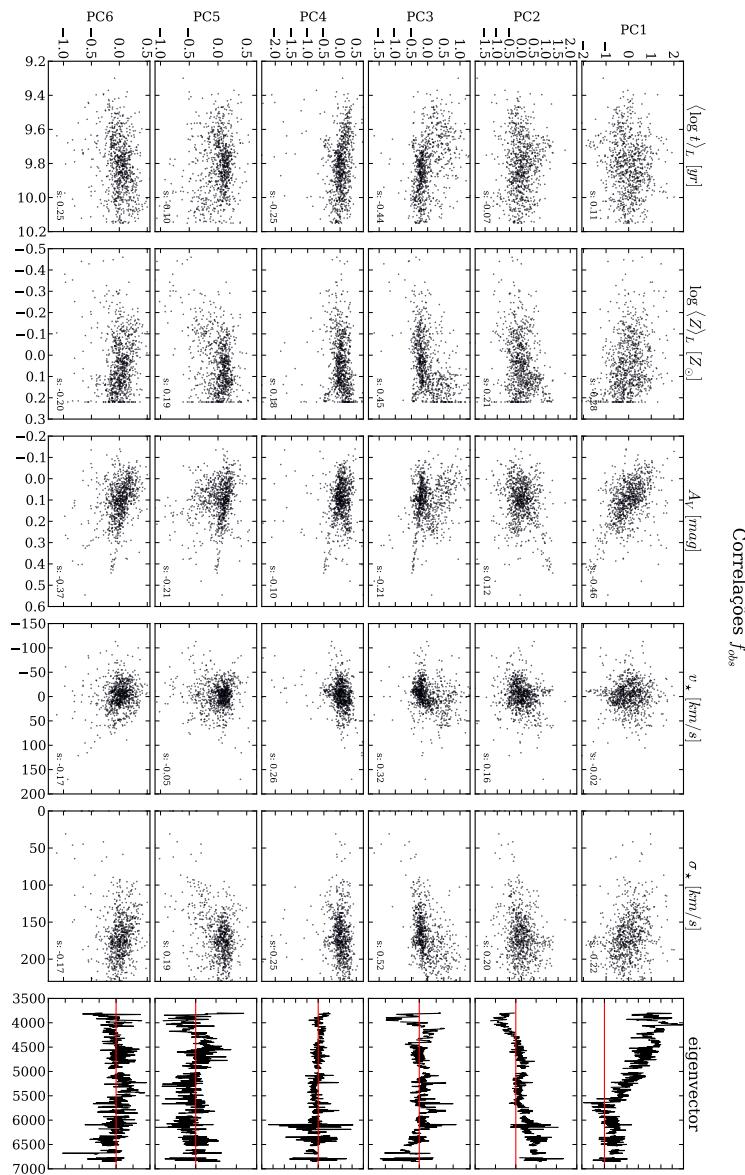


Figura 5.29: Igual a Figura 5.5 para a galáxia NGC 6515.

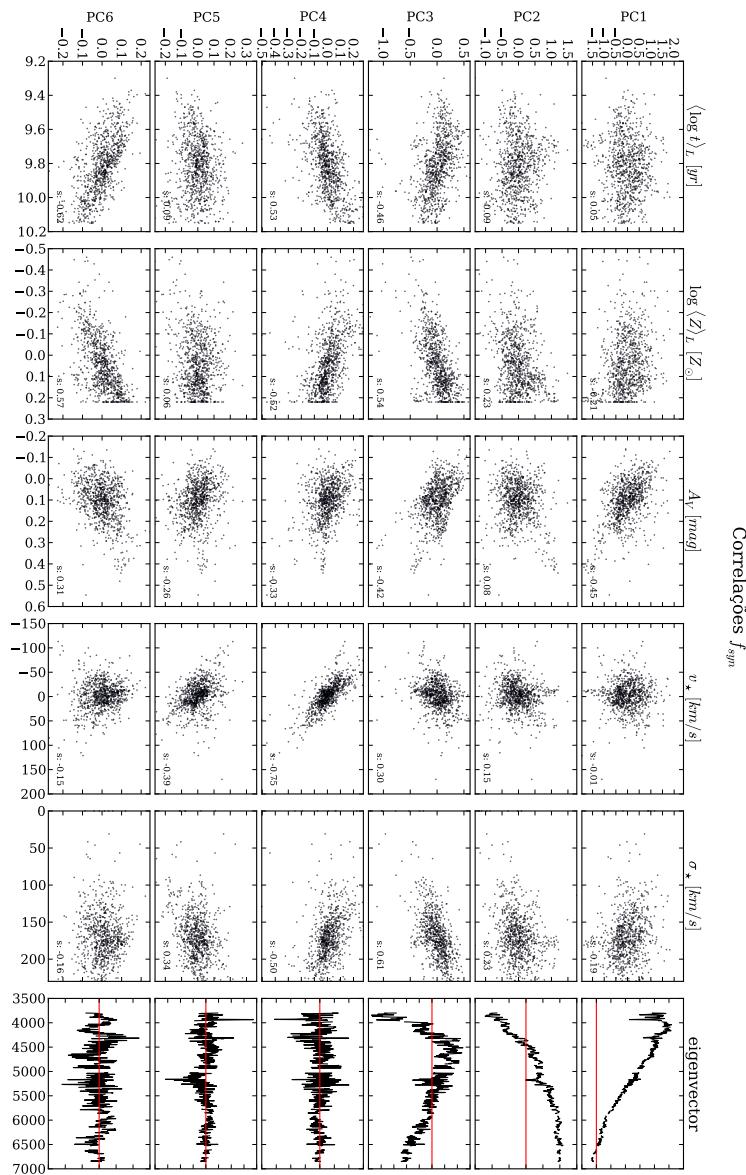


Figura 5.30: Igual a Figura 5.6 para a galáxia NGC 6515.

5.4 *Mergers*

Se para galáxias *early-type* esperávamos (e observamos) uma certa homogeneidade nas distribuições espaciais de PCs e propriedades físicas, o contrário acontece no caso de *mergers*. Estes objetos complexos representam misturas de duas galáxias, às quais se adiciona toda "confusão" gerada pela concentração de gás e poeira e a formação estelar resultante. Nessa seção analisaremos dois sistemas desse tipo.

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 espacial 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 (típica de estrelas A, e portanto de populações de 0.1–1 Gyr) 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 no 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).

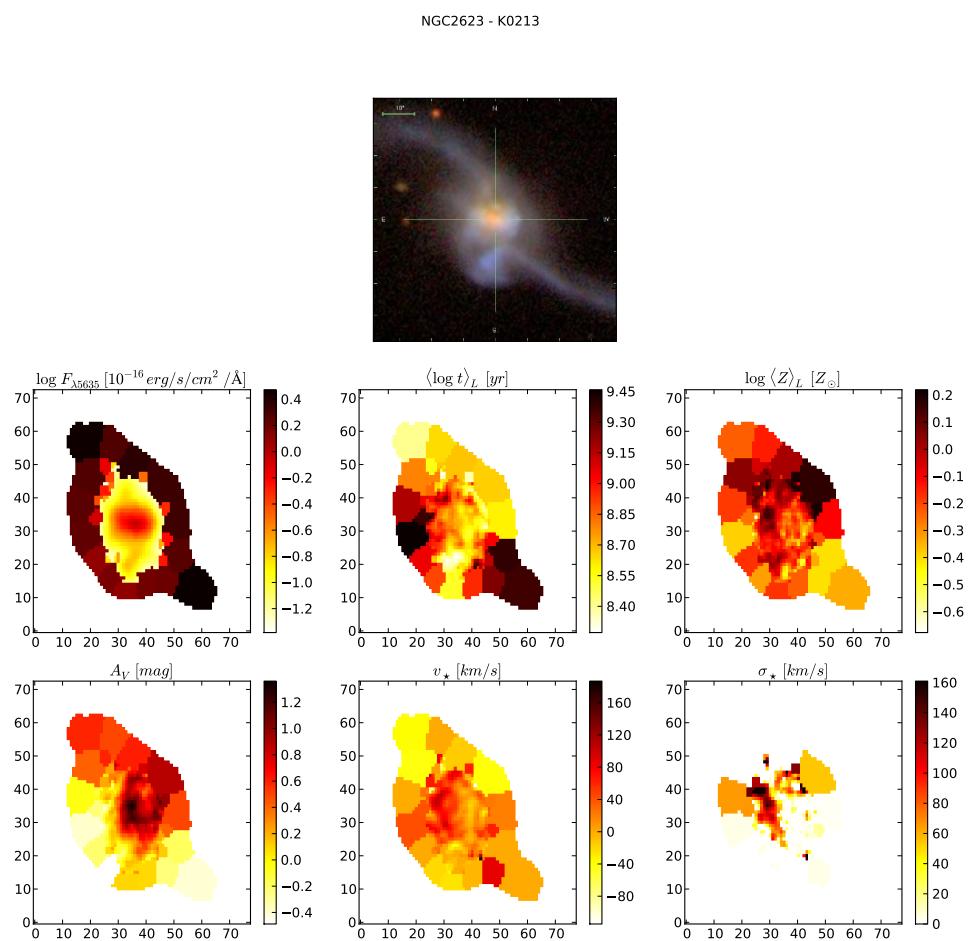


Figura 5.31: Igual a Figura 5.1 para a galáxia NGC 2623.



Figura 5.32: Igual a Figura 5.2 para a galáxia NGC 2623.

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. Para o caso observado a PC3 correlaciona melhor com a velocidade estelar. No caso sintético vemos vestígios da sequência 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) parace ter correlação com nenhum dos parâmetros físicos comparados. O mesmo para a PC6 em ambos os casos.

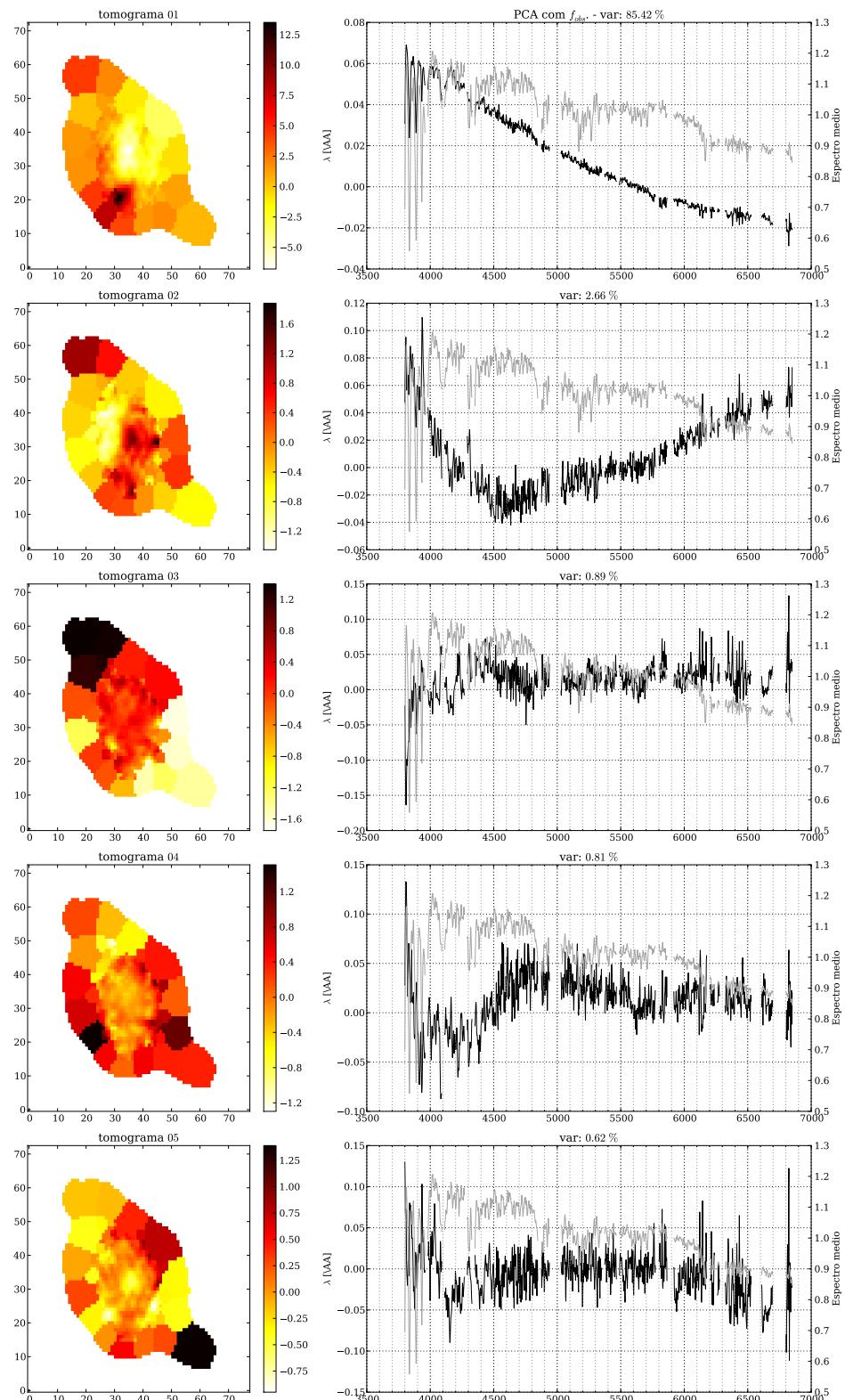


Figura 5.33: Igual a Figura 5.3 para a galáxia NGC 2623.

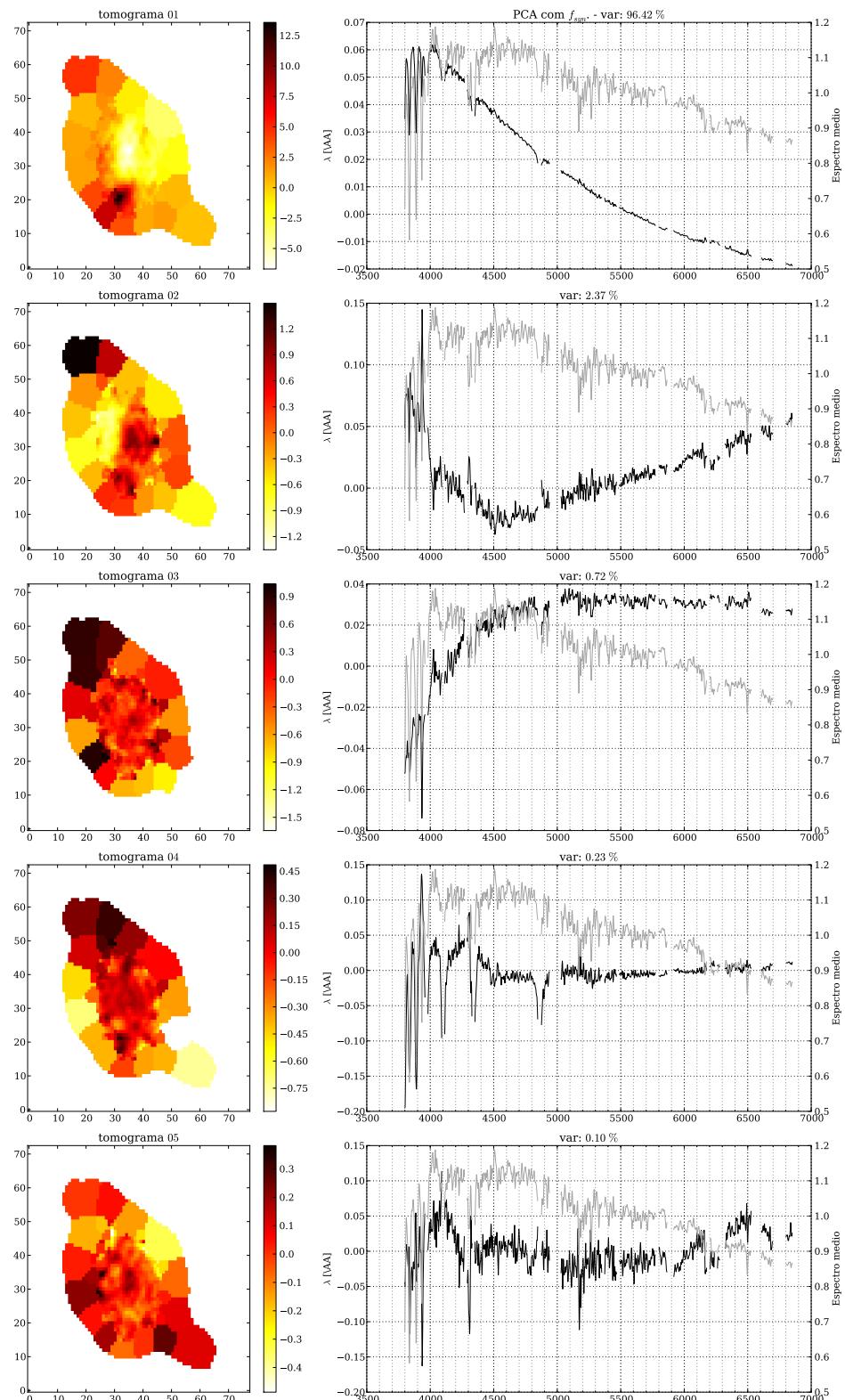


Figura 5.34: Igual a Figura 5.4 para a galáxia NGC 2623.

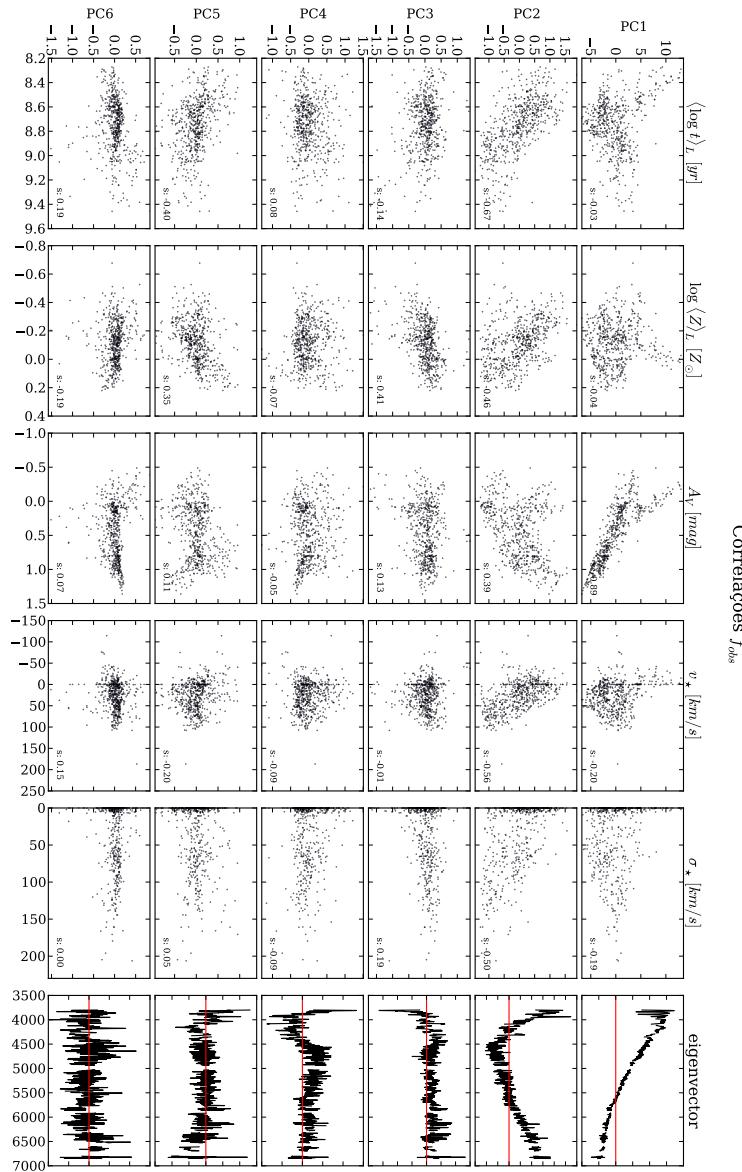


Figura 5.35: Igual a Figura 5.5 para a galáxia NGC 2623.

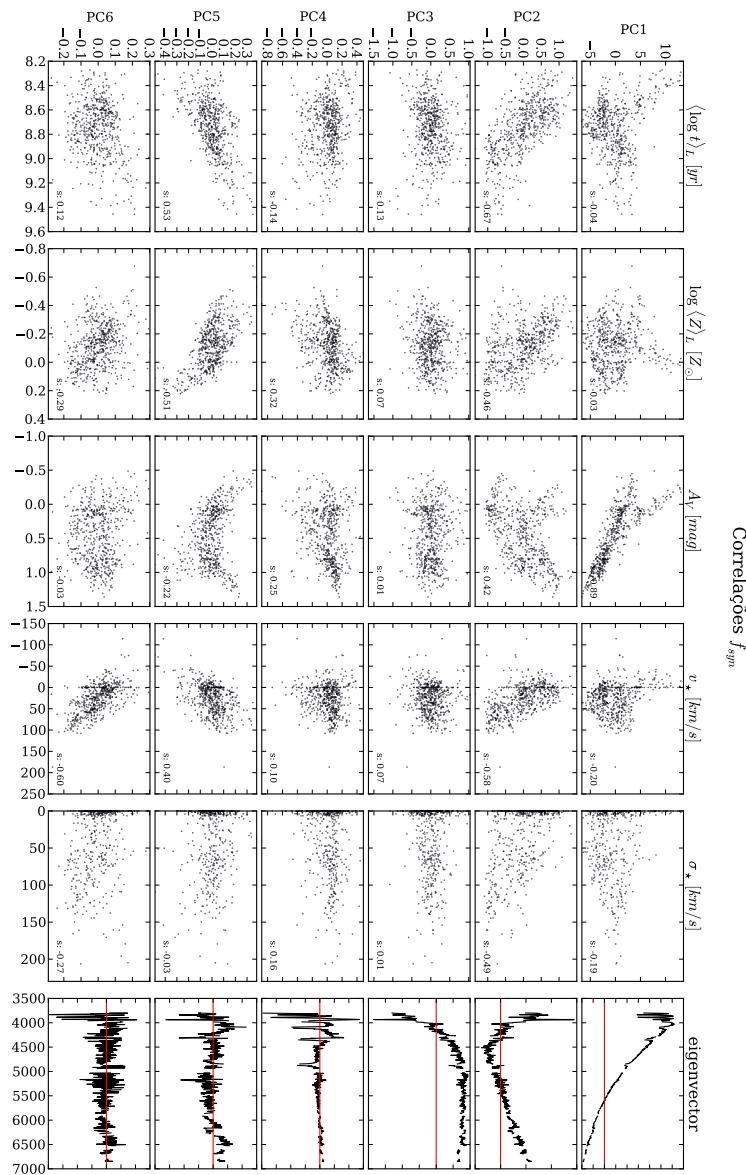


Figura 5.36: Igual a Figura 5.6 para a galáxia NGC 2623.

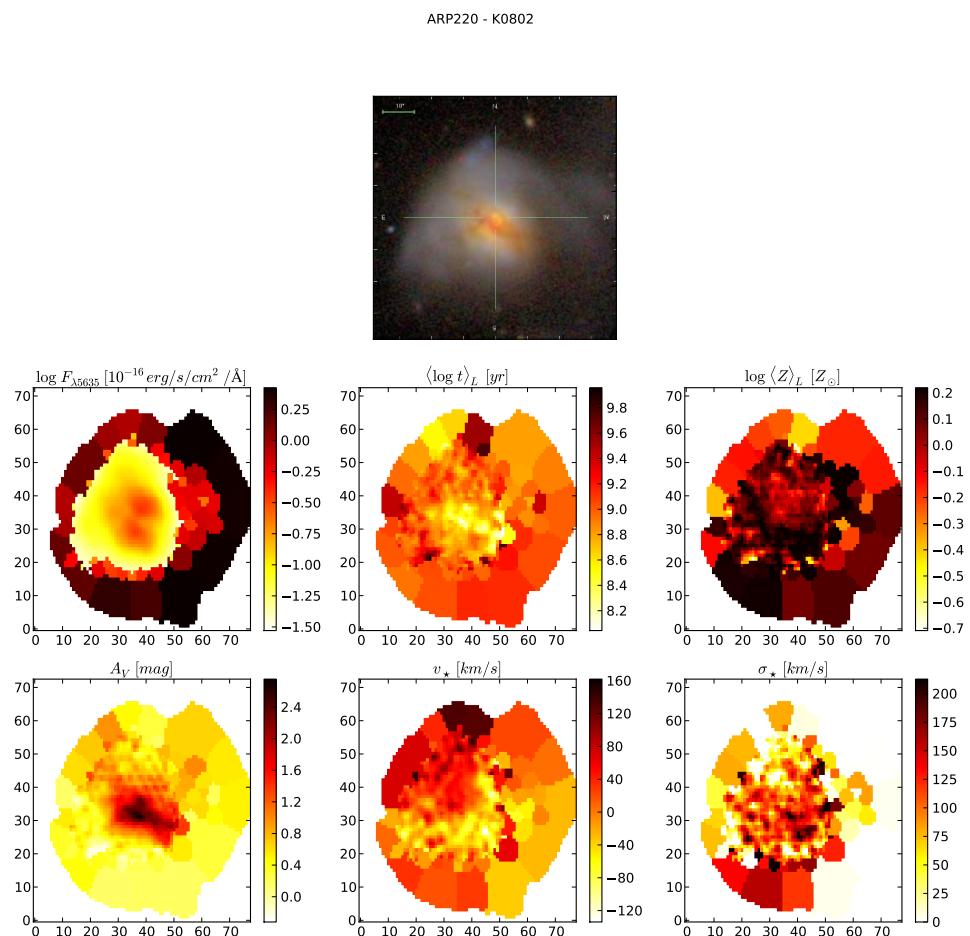


Figura 5.37: Igual a Figura 5.1 para a galáxia ARP 220.

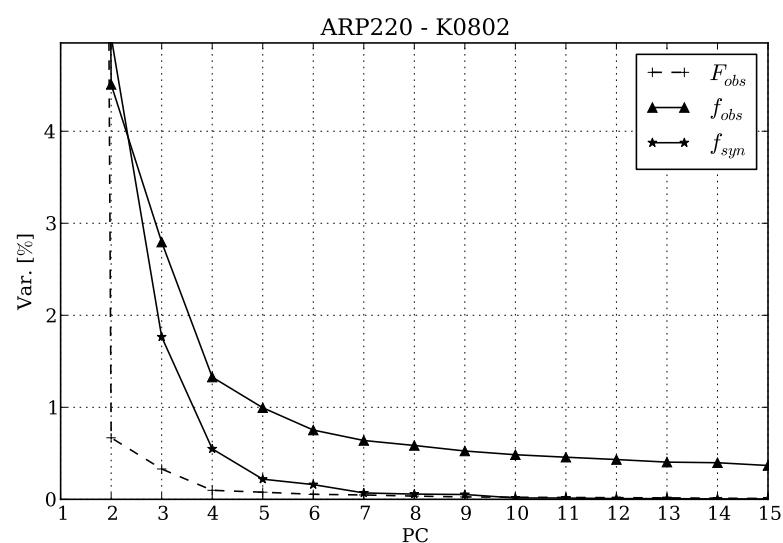


Figura 5.38: Igual a Figura 5.2 para a galáxia ARP 220.

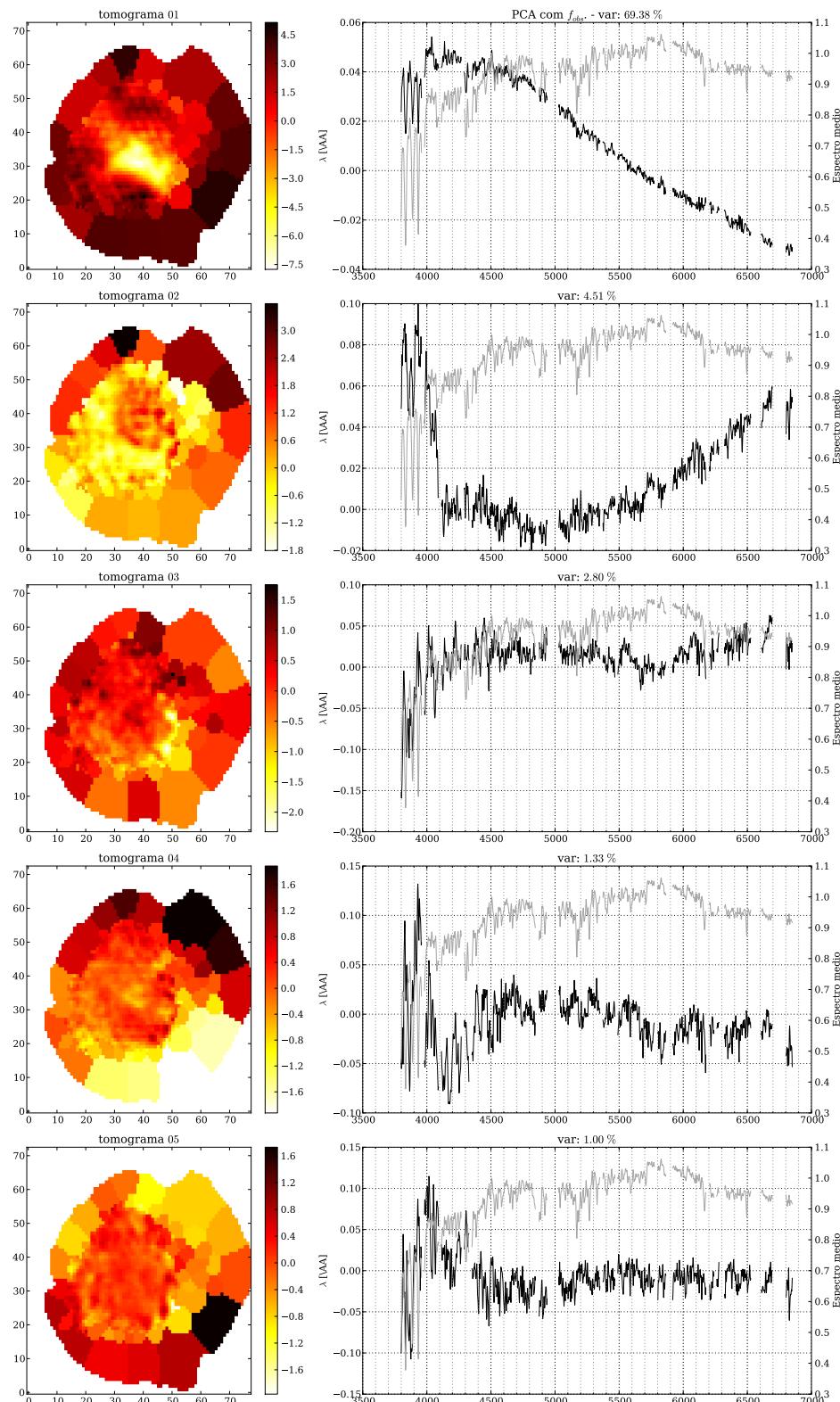


Figura 5.39: Igual a Figura 5.3 para a galáxia ARP 220.

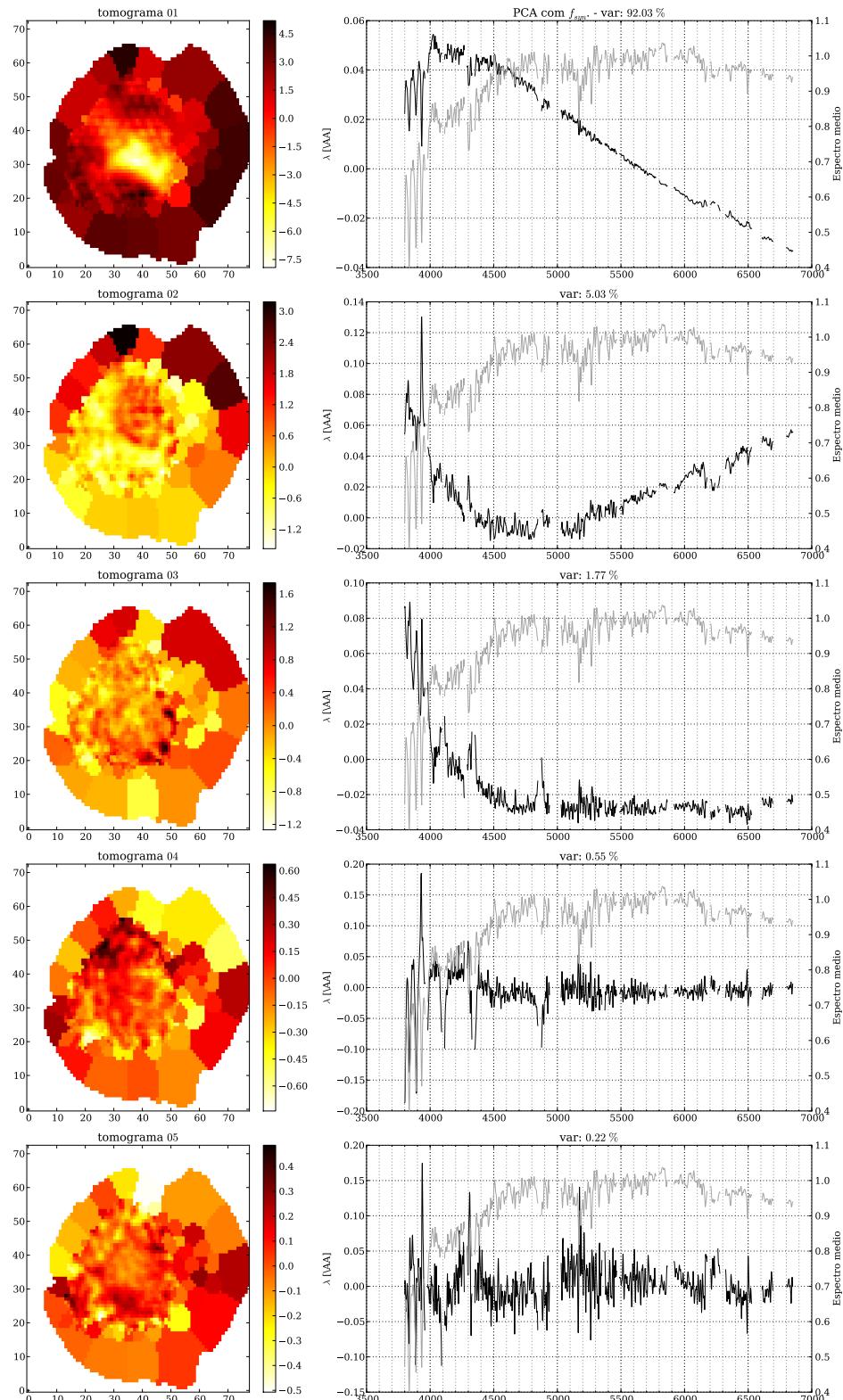


Figura 5.40: Igual a Figura 5.4 para a galáxia ARP 220.

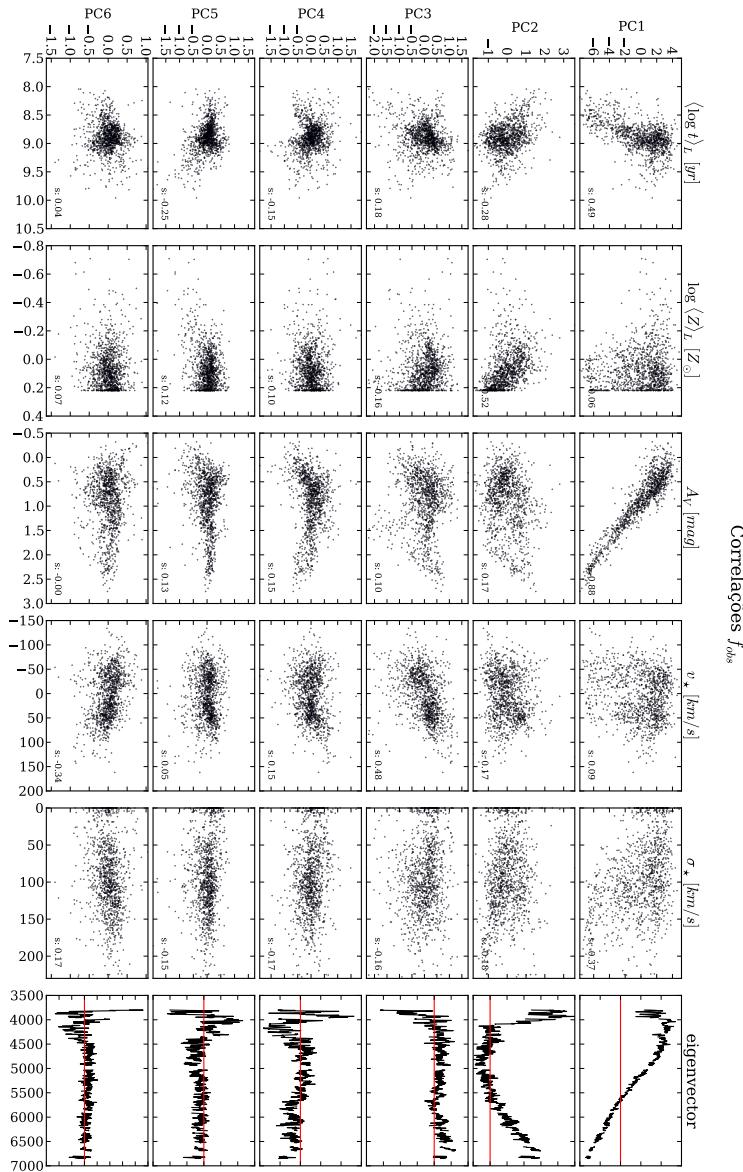


Figura 5.41: Igual a Figura 5.5 para a galáxia ARP 220.

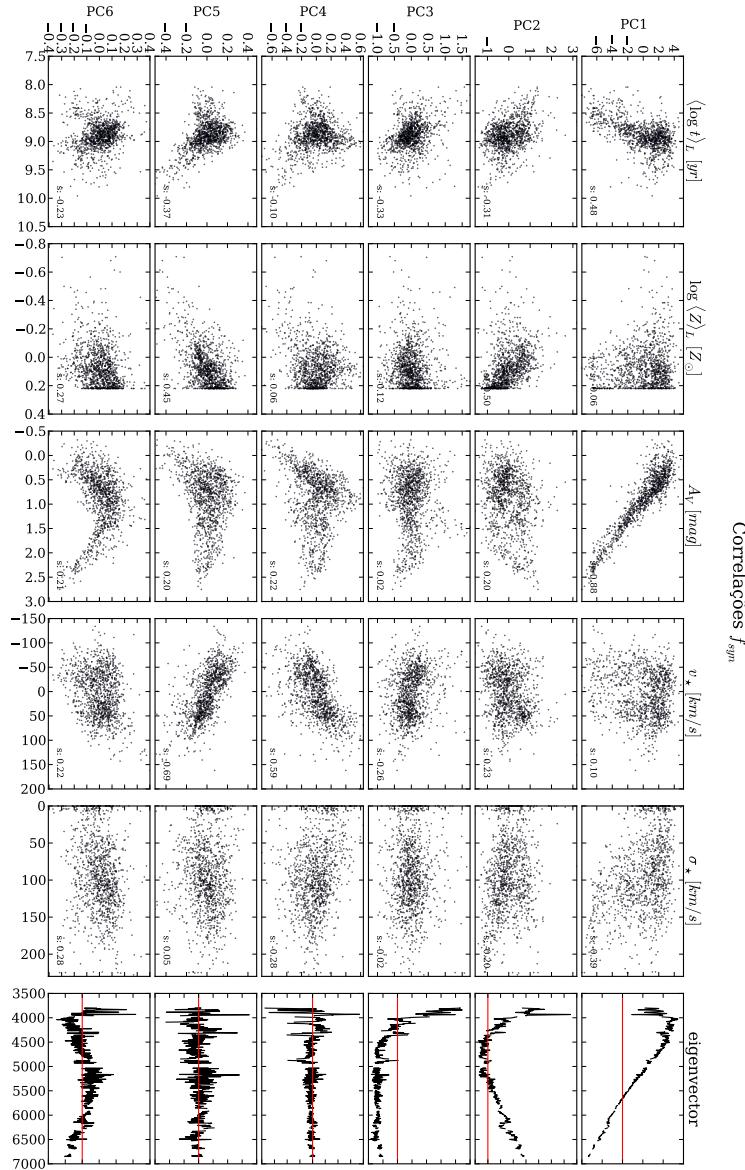


Figura 5.42: Igual a Figura 5.6 para a galáxia ARP 220.

5.5 Considerações gerais

Nesse capítulo exploramos algumas galáxias do CALIFA utilizando a Tomografia PCA para os espectros, observados e sintéticos, normalizados. Fomos capazes de encontrar diversas características nas componentes, sejam correlacionadas com efeitos físicos ou com problemas nos espectros originais.

Os *scree tests* mostram um comportamento assintótico semelhante em todos os casos, com a informação em variância mais comprimida nas primeiras PCs para o caso F_{obs} , seguido por f_{syn} e f_{obs} . Cabe aqui ressaltar que mesmo com o mesmo comportamento assintótico, o número de componentes que “explicam” cada galáxia difere muito. Podemos ver isso na Tabela 5.2 que apresenta a variância acumulada para as 6 primeiras componentes (Λ_6) e o número de componentes necessárias para obtermos 90% da variância acumulada.

Tabela 5.2: Tabela contendo a variância acumulada pelas 6 primeiras PCs (Λ_6), o número de componentes necessárias para cobrir 90% da variância acumulada ($k_{90\%}$) e o valor total da soma de todas os autovalores (Λ). Todos os valores são calculados para os casos observado e sintético, ambos normalizados.

Nome da galáxia	CALIFA ID	Λ_6^{obs} [%]	Λ_6^{syn} [%]	$k_{90\%}^{obs}$	$k_{90\%}^{syn}$	Λ^{obs}	Λ^{syn}
NGC 2916	K0277	89.01	99.81	12	1	14.24	12.43
NGC 0001	K0008	59.61	99.15	82	2	5.33	2.69
NGC 0776	K0073	66.24	99.53	95	1	8.07	4.76
NGC 4210	K0518	85.90	99.80	37	1	10.30	7.79
NGC 1167	K0119	45.79	98.47	161	3	3.74	1.04
NGC 6515	K0864	40.31	98.80	104	3	2.28	0.59
NGC 2623	K0213	90.79	99.90	5	1	15.47	13.64
ARP 220	K0802	79.76	99.75	40	1	10.98	8.21

Muita informação contida nessa tabela ainda necessita ser explorada, mas podemos ver que a soma dos autovalores das PCs para cada galáxia (Λ), que é igual a soma da variância de todos os comprimentos de onda através das zonas, parecem crescer juntamente com a “complexidade” da galáxia analisada. A variância acumulada para os casos observado e sintético parecem confirmar que a informação fica mais comprimida para as primeiras PCs para o segundo caso, em relação ao primeiro.

Pelo fato de que o campo do CALIFA engloba $\sim 100\%$ da luz da galáxia, as estruturas mapeadas pelos cubos de espectros são diversas, mas não nos impediram de encontrar certas similaridades nos resultados para diferentes galáxias com o mesmo tipo morfológico. As galáxias espirais possuem geralmente a PC1 correlacionada com a idade, diferentemente dos *mergers*, que geralmente parecem ter sua componente principal fortemente ligada à extinção estelar (A_V). Encontramos também correlação entre A_V e a primeira componente das galáxias *early-type*. O fato de que o STARLIGHT encontre A_V alto nessas galáxias e que a PC1 correlacione com esse A_V é surpreendente, pois não se espera uma quantidade significativa de poeira em *early-types*. Experimentos com a nova versão do *pipeline* de redução (v1.4) mostram que os parâmetros de extinção são menores e as idades levemente maiores. Acreditamos, portanto, que essa pode ser a causa desse resultado não esperado, mas isso ainda deve ser investigado com mais cuidado. Propriedades cinemáticas aparecem de forma singular, ou composta por mais de uma componente, em todos os casos para todas as galáxias. Já as correlações com a metalicidade aparecem geralmente misturadas com idade ou extinção.

Cabe aqui também ressaltar que através da Tomografia PCA dessas galáxias pudemos encontrar problemas em alguns espectros das galáxias NGC 0001 e NGC 4210. Na última delas o problema parece ser um pouco mais sério e mais difícil de se contornar, mas para a galáxia NGC 0001 podemos realizar novamente a PCA cortando 40 Å do lado azul e 10 Å do lado vermelho do espectro. Talvez para o caso da NGC 4210 possamos reconstruir os cubos de espectros originais suprimindo determinadas componentes da forma que foi feita por S09 ao reconstruir o espectro do AGN da galáxia NGC 4736 e também proposto na Seção 6.2.

Capítulo 6

Conclusões e perspectivas

6.1 Este trabalho

O objetivo principal deste trabalho foi explorar os cubos de espectros das galáxias do *survey* CALIFA utilizando a técnica de Tomografia PCA. Talvez pela primeira vez em IFS de galáxias inteiras, mas com certeza pela primeira com os dados deste *survey*, que cobrem $\sim 100\%$ da luz de uma galáxia, reamostrada em spaxels de 1 arcsec^2 , formando um cubo de espectros $\mathbf{F}_{x,y,\lambda}^{orig}$.

Ao longo dos últimos anos a PCA vem se tornando uma técnica quase obrigatória para o estudo de conjuntos multivariados de dados. Através dessa redução de dimensionalidade pressupõe-se que a busca pelas características mais importantes sejam mapeadas pelas componentes principais. A descoberta de linhas largas (elementos típicos de AGNs tipo 1) na galáxia NGC 4736 classificada como LINER, através da técnica de tomografia PCA desenvolvida no trabalho de S09, usando dados de IFS do Gemini da região central da galáxia ($\sim 100 \text{ pc}$) corrobora essas pressuposições, servindo de base teórica para nosso trabalho. Cabe aqui ressaltar que o *FoV* coberto pelos dados do Gemini usados no trabalho de S09 equivale a ~ 2 elementos de resolução do CALIFA. Com um campo cobrindo toda galáxia, nossos dados mapeiam regiões com estruturas muito diferentes (e.g., bojo, disco, regiões HII).

Desenvolvemos um programa, por ora apelidado de “PCAlifa”, utilizando linguagem Python e os *pipelines* QBICK e PyCASSO, ambos construídos pela colaboração entre o projeto CALIFA e o GAS-UFSC, que serviu de base para todo nosso estudo exploratório. Estudo este sobre 4 galáxias espirais: NGC 0001 (CALIFA 8), NGC 0776 (CALIFA 73), NGC 2916 (CALIFA 277) e NGC 4210 (CALIFA 518); 2 galáxias *early-type*: NGC 1167 (CALIFA 119) e NGC 6515 (CALIFA 864); 2 *mergers*: NGC 2623 (CALIFA 213) e ARP 220 (CALIFA 802). Em

duas das galáxias escolhidas pudemos, através da Tomografia PCA, identificar problemas com os espectros que passaram para o cubo final mesmo com os tratamentos dos *pipelines* de redução e os filtros de qualidade do QBICK.

Encontramos diversas correlações entre as primeiras PCs (autoespectros) e propriedades físicas de populações estelares obtidos através da síntese para os cubos de espectros das galáxias do CALIFA, feito por CF13 utilizando o STARLIGHT. Essas correlações formam uma certa engenharia reversa no sentido de buscar os parâmetros mais “importantes” em variância (as PCs) que mapeiam propriedades físicas, de uma forma não-paramétrica.

6.2 Trabalhos futuros

Cumprimos um papel exploratório inicial utilizando os dados de IFS do CALIFA. Temos diversas vertentes surgindo a partir deste trabalho, mas a primeira delas é poder fazer a mesma análise utilizando a nova versão do *pipeline* de redução do CALIFA (v1.4). Essa nova versão ficou pronta há poucos meses e apenas em Janeiro foi testada por nossos colaboradores. Através desses testes preliminares já pôde-se perceber que esta nova redução produz espectros com melhor calibração.

Em todo esse estudo trabalhamos com os cubos rearranjados em zonas de Voronoi. Essa escolha se baseia única e exclusivamente em nosso desejo de comparar as PCs com os resultados do STARLIGHT, que requer um S/N grande o suficiente para obter resultados fiáveis. Porém, esse agrupamento acaba por deteriorar muito a qualidade das imagens para zonas externas, que, quanto mais afastadas do núcleo, possuem mais pixels agrupados em uma zona, perdendo assim resolução espacial. Seria interessante fazer o PCA com o cubo original, não zonificado. Isto talvez produza tomogramas que revelem melhor as estruturas como braços espirais e regiões HII nas partes externas das galáxias.

Um outra possível linha de trabalho é poder explorar as qualidades de filtro da PCA, seja para reconstruir os espectros originais selecionando determinadas características encontradas em diferentes PCs ou para suavização dos espectros através da redução do ruído causado por todo processo observacional, de redução e de criação do COMBO. De posse da matriz de autoespectros ($\mathbf{E}_{\lambda k}$) podemos reconstruir a matriz $\mathbf{I}_{z\lambda}$ calculada pela equação 3.3 utilizando $r \leq k$ autoespectros, onde k é o número total de autoespectros, através da transformação

$$\mathbf{I}_{z\lambda}^{rec}(r \leq k) = \mathbf{T}_{zk}(r \leq k) \cdot [\mathbf{E}_{\lambda k}(r \leq k)]^T \quad (6.1)$$

assim nos permitindo reconstruir o cubo de espectros ($\mathbf{F}_{z\lambda}^{rec}$) adicionando o espectro médio $\langle \mathbf{F}_\lambda \rangle$ a $\mathbf{I}_{z\lambda}^{rec}$:

$$\mathbf{F}_{z\lambda}^{rec} = \mathbf{I}_{z\lambda}^{rec} + \langle \mathbf{F}_\lambda \rangle \quad (6.2)$$

Por exemplo, S09 reconstroem o espectro do AGN da galáxia NGC 4736 selecionado determinadas componentes. Outro exemplo é a reconstrução dos espectros com a supressão de componentes que caracterizem cinemática (PCs que correlacionam com v_\star e σ_\star). Outras técnicas podem ser aliadas ao PCA, como, por exemplo, faz ?) utilizando Tomografia PCA juntamente com uma técnica de *wavelet* com o intuito de filtragem do cubo de espectros.

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. 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.

Um outro experimento que podemos realizar é fazer a PCA utilizando como entrada os logarítmos dos espectros. Essa variante da análise pode ser vista como uma mera curiosidade matemática, uma experiência para ver o que se obtém. No entanto, ela é inspirada em um motivo físico: é razoável esperar que utilizando o logaritmo possamos ressaltar efeitos devidos a gradientes de poeira, já que a extinção atua de forma multiplicativa sobre os espectros.

Devido a característica de *legacy-survey* do CALIFA e do fato que cada galáxia é uma amostra estatística *per se*, muitas novidades ainda virão através da ciência realizada com estes dados pelo grupo de pesquisadores do IAA e também através da parceria com nossos colaboradores aqui do GAS-UFSC.