

# **!OJO! Tomografia PCA aplicado às galáxias do Projeto CALIFA Survey**

Eduardo Alberto Duarte Lacerda

Orientador:

Prof. Dr. Roberto Cid Fernandes Jr.

• • •

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

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

Florianópolis (SC) – 14 de fevereiro de 2014

# Sumário

|          |                                                                             |           |
|----------|-----------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Introdução</b>                                                           | <b>1</b>  |
| 1.1      | Um labirinto de dados . . . . .                                             | 1         |
| 1.2      | A nova ferramenta - Tomografia PCA . . . . .                                | 2         |
| 1.3      | Este trabalho . . . . .                                                     | 3         |
| 1.3.1    | Organização deste trabalho . . . . .                                        | 4         |
| <b>2</b> | <b>O projeto CALIFA e o <i>pipeline</i> PyCASSO</b>                         | <b>6</b>  |
| 2.1      | O survey CALIFA . . . . .                                                   | 6         |
| 2.1.1    | A “ <i>colmeia</i> ” de fibras - os dados do CALIFA usados neste trabalho . | 7         |
| 2.2      | O <i>pipeline</i> PyCASSO . . . . .                                         | 9         |
| 2.2.1    | E/S de dados no PyCASSO . . . . .                                           | 9         |
| 2.2.2    | Exemplos de utilização . . . . .                                            | 10        |
| <b>3</b> | <b>PCA e Tomografia PCA</b>                                                 | <b>14</b> |
| 3.1      | Principal Component Analisys . . . . .                                      | 14        |
| 3.1.1    | PCA das galáxias do CALIFA . . . . .                                        | 15        |
| 3.2      | Tomografia PCA . . . . .                                                    | 17        |
| 3.2.1    | Evidências de linhas largas . . . . .                                       | 17        |
| <b>4</b> | <b>PCA em IFS - (different flavors?!?! FIXME PLEASE)</b>                    | <b>21</b> |

|                                   |                                                                         |           |
|-----------------------------------|-------------------------------------------------------------------------|-----------|
| 4.1                               | Normalização . . . . .                                                  | 21        |
| 4.2                               | Fluxos observados e sintéticos . . . . .                                | 26        |
| 4.3                               | Comparando as PCs com o <i>STARLIGHT</i> : engenharia reversa . . . . . | 30        |
| <b>5</b>                          | <b>Resultados</b>                                                       | <b>39</b> |
| <b>6</b>                          | <b>Conclusões e perspectivas</b>                                        | <b>40</b> |
| 6.1                               | Este trabalho . . . . .                                                 | 40        |
| 6.2                               | Trabalhos futuros . . . . .                                             | 40        |
| 6.2.1                             | Linhas de emissão e intervalos específicos em comprimento de onda .     | 40        |
| <b>A</b>                          | <b>[FIXME] Python</b>                                                   | <b>41</b> |
| <b>Referências Bibliográficas</b> |                                                                         | <b>I</b>  |

# **Lista de Figuras**

|     |                                                                        |    |
|-----|------------------------------------------------------------------------|----|
| 2.1 | Diagrama cor-magnitude para as galáxias do CALIFA.                     | 8  |
| 2.2 | Configuração do <i>bundle</i> de fibras do PPMAS/PPAK.                 | 8  |
| 2.3 | Exemplo de programa utilizando PyCASSO                                 | 11 |
| 2.4 | Programa idade estelar média                                           | 11 |
| 2.5 | Mapa da idade estelar média da galáxia NGC 2916                        | 12 |
| 2.6 | Exemplo de programa para perfil radial                                 | 12 |
| 2.7 | Perfil radial da idade estelar média da galáxia NGC 2916               | 13 |
| 3.1 | Exemplo de cálculo de PCA usando o PyCASSO e SciPy                     | 16 |
| 3.2 | <i>Scree test</i> na galáxia NGC 4736.                                 | 18 |
| 3.3 | Tomograma e autoespectro 1 da galáxia NGC 4736.                        | 19 |
| 3.4 | Tomograma e autoespectro 2 da galáxia NGC 4736.                        | 19 |
| 3.5 | Tomograma e autoespectro 3 da galáxia NGC 4736.                        | 20 |
| 4.1 | Tomogramas de 1 a 5 da galáxia NGC 2916 - $F_{obs}$ .                  | 24 |
| 4.2 | Tomogramas de 1 a 5 da galáxia NGC 2916 - $F_{obs}/F_{\lambda 5365}$ . | 25 |
| 4.3 | Fluxos de normalização para cada zona da galáxia K0277.                | 26 |
| 4.4 | Scree test comparativo entre 3 PCAs.                                   | 27 |
| 4.5 | Tomogramas de 1 a 5 da galáxia NGC 2916 - $F_{syn}/F_{\lambda 5365}$ . | 29 |

|      |                                                          |    |
|------|----------------------------------------------------------|----|
| 4.6  | Correlações PCs vs. parâmetros físicos - $F_{obs}$ norm. | 31 |
| 4.7  | Correlações PCs vs. parâmetros físicos - $F_{syn}$ norm. | 32 |
| 4.8  | Dados no espaço das PCs vs AV- $F_{obs}$ norm.           | 33 |
| 4.9  | Dados no espaço das PCs vs AV- $F_{syn}$ norm.           | 34 |
| 4.10 | Correlações PCs vs. parâmetros físicos - $F_{obs}$ .     | 35 |
| 4.11 | Exemplo de máscaras em um espectro do cubo de dados.     | 37 |

## **Lista de Tabelas**

|                                                                       |    |
|-----------------------------------------------------------------------|----|
| 2.1 Relação de pixels e zonas em algumas galáxias do CALIFA . . . . . | 10 |
|-----------------------------------------------------------------------|----|

# Capítulo 1

## Introdução

### 1.1 Um labirinto de dados

Cientistas hoje, em sua maioria, encontram-se perdidos em meio a um labirinto de informações. Essas são apenas parte de um cenário incompleto, mas não falso, que chamamos de nosso universo. Com o avanço tecnológico, melhores formas de se obter informações tornam cada vez mais evidente o surgimento de um *fordismo*<sup>1</sup> relativo a informações que auxiliam na criação desse labirinto, mas ao mesmo tempo fomenta a criatividade e curiosidade de cientistas que se tornam aventureiros na busca da saída. Mas essa fuga é apenas a primeira etapa assim se tornando parte basilar na formação da pesquisa científica.

Os primeiros *surveys*<sup>2</sup> astronômicos surgem com a inata curiosidade do homem de observar tudo a sua volta e registrar suas observações. O crescimento desses catálogos é produto direto da evolução dos equipamentos usados nestas investigações. No início da década de 80 (Huchra et al. 1983; Huchra 1988; da Costa et al. 1988) a astronomia extragalática entra nesse cenário de produção sistemática em massa de dados. De lá para cá a quantidade de dados só aumenta e, com a criação dos *mega-surveys* (*SDSS*; York et al. 2000) (2dFGRS; Colless 1999) (2MASS; Skrutskie et al. 2006), alguns já terminados, outros ainda começando ou por terminar (LSST; Ivezić et al. 2008) (J-PAS; Benítez et al. 2009), estamos à beira do humanamente impossível de se avaliar, necessitando assim a ajuda de máquinas e métodos computacionais cada vez mais eficientes.

---

<sup>1</sup>O termo *fordismo* foi criado por Antonio Gramsci em 1922 e é relativo a Henry Ford e o surgimento da produção em massa de automóveis no início do século XX.

<sup>2</sup>Um *survey* astronômico é um levantamento de informações ou mapeamento de regiões do céu utilizando telescópios e detetores.

Com esse crescimento exponencial na quantidade de dados, precisamos cada vez mais de ferramentas matemáticas/estatísticas. No contexto de espectros de galáxias, uma ferramenta muito útil para síntese espectral é o STARLIGHT, desenvolvido por Cid Fernandes et al. (2005). Hoje, com o uso de painéis de fibras óticas apontadas para as galáxias temos os *surveys* de IFS (*Integral Field Spectroscopy*) onde passamos a obter dezenas, centenas e até milhares de espectros por galáxia, obtendo então dimensão espacial também nos dados. Assim temos para cada píxel (duas dimensões espaciais) um espectro (uma dimensão espectral), formando assim um cubo de *spaxels*<sup>3</sup>. O pioneiro nessa produção em massa de dados é o *Calar Alto Legacy Integral Field spectroscopy Area survey*<sup>4</sup> (CALIFA; Sánchez et al. 2012), produzindo cerca de mil espectros por galáxia observada. Outros *mega-surveys* IFS estão por vir (veja a seção 2.1).

Com esses cubos de dados em mão, podemos assim executar o STARLIGHT para cada *spaxel* e então obtemos propriedades físicas em função da posição na galáxia (Cid Fernandes et al. 2013). Uma ferramenta menos astrofísica, mas não menos utilizada, é o PCA (*Principal Component Analysis*) e no presente trabalho fazemos seu uso juntamente com uma técnica criada por Steiner et al. (2009) chamada de Tomografia PCA, onde une-se imageamento e espectrografia ao mesmo tempo.

## 1.2 A nova ferramenta - Tomografia PCA

A técnica de Análise de Componentes Principais (PCA) é simples, não-paramétrica e nos ajuda a extrair informações de conjuntos de dados com muitas variáveis, reduzindo a dimensionalidade no sentido de encontrar quais são os elementos com maior variâncias no seu conjunto de dados. No caso de um espectro, temos uma medida de fluxo para cada intervalo em comprimento de onda. Em um cubo de spaxels, temos as dimensões espaciais também, obtendo assim uma infinidade de variáveis. O grande problema do PCA é que você tem a resposta, mas não sabe a pergunta.

Hoje, PCA é utilizado exaustivamente em várias áreas de conhecimento, principalmente em reconhecimento de padrões, computação visual e filtragem e compactação de dados (Kamruzzaman et al. 2010; Borcea et al. 2012). Podemos ver exemplos também em medicina (Balakrishnan et al. 2013). Na astrofísica o PCA passeia por diversas ramificações da área. A luz de um objeto até nossos telescópios sofre influência de muito ruído, devido a inúmeros

---

<sup>3</sup>spectral pixels

<sup>4</sup><http://www.caha.es/CALIFA/>

problemas como: atenuações, avermelhamento por poeira, contaminação ótica através de objetos que estejam no mesmo FoV, entre outros; a diversa quantidade de instrumentos e suas complexidades também geram diversas assinaturas indesejadas; por todos esses motivos fica claro que os dados geralmente necessitam de uma boa filtragem. Junto com outras técnicas (tomogramas, wavelets, Fourier), o PCA vem sendo muito utilizado para filtragem de dados, principalmente cubos de dados advindos de IFS que possuem muitas dessas assinaturas instrumentais (Riffel et al. 2011). Outro exemplo de uso de PCA aparece no artigo IV do grupo SEAGal/STARLIGHT (Mateus et al. 2007) auxiliando no estudo da dependência ambiental de algumas propriedades físicas (idade estelar média ponderada pela luz, massa estelar, metalicidade estelar e *mass-to-light ratio ( $M_{\star}/L$ )*) em uma amostra de galáxias do SDSS DR4. Em Chen et al. (2012) é criado 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. Usando uma minimização quadrática encontram quais os coeficientes e qual o número ótimo de PCs que melhor estimam os parâmetros físicos dos espectros modelo. Então projetam os espectros observados pelo SDSS DR7 (Abazajian et al. 2009) e pelo Baryon Oscillation Spectroscopic Survey (BOSS Ahn et al. 2012) para galáxias massivas com  $z \sim 0.6$  até o presente, atribuindo um sentido físico a cada PC.

Com a criação da técnica de Tomografia PCA citada anteriormente 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, aliando a infinidade de informações físicas presentes nos espectros e nas imagens. Além do artigo de Steiner et al. (2009), outros exemplos de uso de Tomografia PCA podem ser encontrados em Riffel et al. (2011) e Ricci et al. (2011).

## 1.3 Este trabalho

Através da colaboração do Grupo de Astrofísica da Universidade Federal de Santa Catarina (GAS-UFSC) com o grupo de pesquisadores do projeto CALIFA temos a oportunidade de trabalhar com os dados de IFS das galáxias observadas por esse projeto, que ainda está em andamento. O seu primeiro *Data Release*<sup>5</sup> (Husemann et al. 2013, DR1) possui 100 objetos e por volta de 400 mil espectros. A previsão é que ao término do projeto serão observadas

---

<sup>5</sup>[http://www.caha.es/CALIFA/public\\_html/?q=content/califa-dr1](http://www.caha.es/CALIFA/public_html/?q=content/califa-dr1)

até 600 objetos. Embora outros surveys de IFS estão completos ou em andamento (ver Seção 2.1), o CALIFA é o que podemos chamar de *estado da arte* em surveys de espectroscopia de campo integrado (IFS).

Dado o grande número de informações sobre cada galáxia necessitamos de um *pipeline* que faça a organização de todos os dados para que cálculos e gráficos de mais variadas dificuldade sejam fácil para que até um programador de nível iniciante possa fazer. André L. de Amorim, colaborador de nosso grupo, juntamente com outros colaboradores de nosso grupo e do projeto CALIFA construiu o PyCASSO (*Python CALIFA STARLIGHT Synthesis Organizer*) (Cid Fernandes et al. 2013, cap. 4) que faz a organização dos dados que vêm do *survey* 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 e acredito que não seria realizável em tempo hábil.

### 1.3.1 Organização deste trabalho

No seguinte capítulo apresenta de forma mais detalhada o *survey* CALIFA e o *pipeline* PyCASSO. Nele também são demonstrados alguns exemplos de utilização e gráficos.

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

A partir do quarto capítulo temos os mais diferentes *tipos*<sup>6</sup> de execução do PCA, juntamente com suas implicações aos resultados das análises. Também são apresentados diversos gráficos demonstrando as diferenças entre a utilização de cada pré-processamento, análise da Tomografia PCA e das PCs, bem como as principais referências na área.

Com todo o arcabouço teórico em mãos, no capítulo cinco temos um estudo de caso para 9 objetos do survey CALIFA, escolhidos por seu tipo morfológico. São 2 galáxias *early-type*, 4 espirais e 2 objetos compostos (mergers). Como forma de comparação com os resultados da síntese de populações estelares executadas pelo STARLIGHT nesses objetos também é feito uma espécie de engenharia reversa através de correlações e comparações.

Por fim, temos as conclusões e perspectivas futuras deste trabalho e da nossa colaboração com o projeto CALIFA no sexto e último capítulo. !obj! Devido ao grande número de

---

<sup>6</sup> !obj! na literatura usa-se *flavors*, não sei se devo usar *sabores* também. tiposs aqui é usado apenas como uma maneira de tipificar diferentes pré-processamentos dos dados antes da execução da análise de componentes principais.

imagens e dados para análise, muitas delas ficaram em anexo.

## Capítulo 2

# O projeto CALIFA e o *pipeline* PyCASSO

As observações do universo modificaram completamente o nosso modo de viver, pensar, compreender-se. Aprendemos a contar os dias, desenvolvemos um sistema de meses, estações do ano, movimentos das marés, entre outras coisas que já são parte do senso comum, mas que um dia foram o estado da arte da ciência. É assim que surge o CALIFA: um projeto que está modificando nossa maneira de ver e pensar as galáxias no nosso universo de forma que entendamos melhor a nossa também.

Com a massiva quantidade de dados obtidos, resultado direto de um projeto de ciência de ponta, vem também a dificuldade da interpretação dos dados. No caso do CALIFA, através da *pipeline* PyCASSO (Cid Fernandes et al. 2013), a programação investigativa se torna simples e ao mesmo tempo robusta, facilitando a construção de todo tipo de resultados, sejam eles !objeto! matemáticos ou físicos.

### 2.1 O survey CALIFA

No sul da Espanha, mais precisamente em *Sierra de Los Filabres* (Andalucía), está situado o germano-espânico *Calar Alto Observatory*. O projeto CALIFA está sendo possível através de observações pelo maior de seus 3 telescópios (3.5m) ao longo de 250 noites. Em comparação com o *SDSS*, o CALIFA terá a mesma ordem de número de espectros para estudo ( $\sim 10^6$ ), mas, apesar de um número 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 muito estreitos, dificultando o legado do survey para outras pesquisas científicas mais abrangentes (SAURON; de Zeeuw et al. 2002, região central de 72 galáxias com  $z < 0.01$ ) (PINGS; Rosales-Ortega et al. 2010, algumas galáxias muito próximas ( $\sim 10$  Mpc) e o estudo atual de 70 (U)LIRGs com  $z < 0.26$ ) (VENGA; Blanc et al. 2010, 30 galáxias espirais). Apesar de ser primariamente construído para o estudo da física bariônica da evolução de galáxias, o CALIFA está projetado para que seu legado seja bem abrangente, possibilitando diversos tipos de estudos em diversas áreas. Outros surveys IFU ainda estão por vir, como SAMI (Croom et al. 2012) e MaNGA<sup>1</sup>

### 2.1.1 A “colméia” de fibras - os dados do CALIFA usados neste trabalho

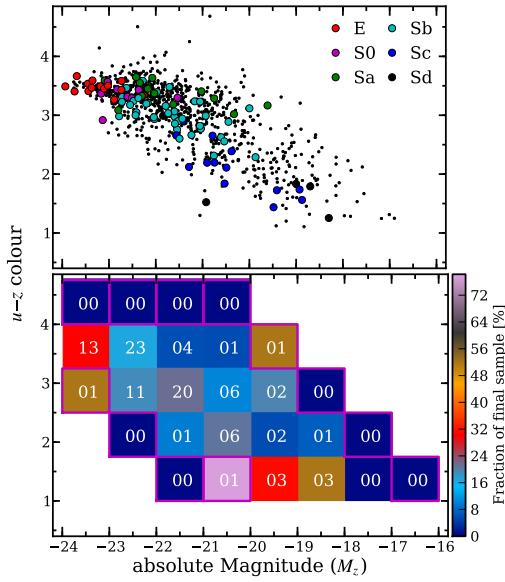
A amostra-mãe do projeto comporta 939 galáxias (das quais  $\sim 600$  serão observadas) com *redshifts* entre  $0.005 < z < 0.03$  que distribuídas cobrem o diagrama cor-magnitude com  $M_r < -18$  (Figura 2.1) em uma ampla variedade de tipos morfológicos, massa em estrelas, condições do gás ionizante.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

Os dados são reduzidos utilizando o programa CALIFA Pipeline versão 1.3c, descrito em Husemann et al. (2013). Os espectros vêm em duas configurações: a V500 cobrindo de  $\sim 3700$  até  $7000$  Å com resolução de  $\sim 6$  Å de largura à meia altura (FWHM) e a V1200 ( $\sim 3650 - 4600$  Å FWHM  $\sim 2.3$  Å). A cobertura do V500 seria 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, criando uma que chamamos de COMBO. A parte com  $\lambda < 4600$  Å vem do V1200 e a outra parte do V500. Os espectros foram reamostrados no mesmo FWHM do V500.

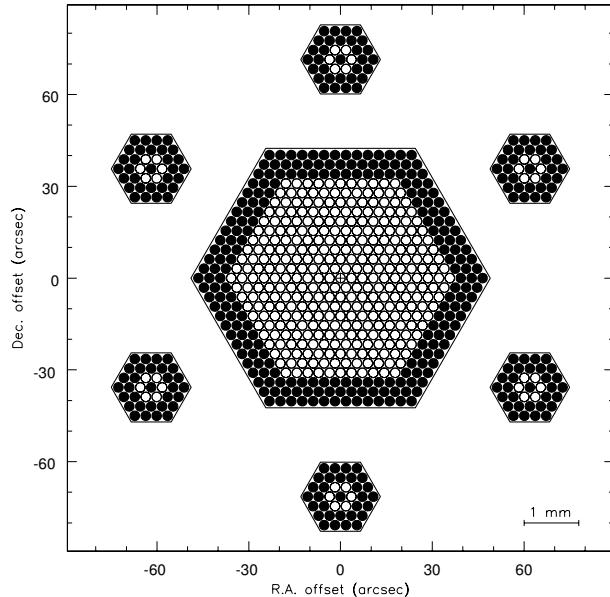
Após os dados passarem pelo CALIFA *Pipeline* v1.3c, os parâmetros físicos aqui usados

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

<sup>2</sup>campo de visão



**Figura 2.1:** Distribuição das galáxias do CALIFA no diagrama  $u - z$  vs.  $M_z$ . *Painel superior*: Em pontos pretos estão as galáxias pertencente a amostra-mãe e em cores as galáxias presente no CALIFA DR1. As diferentes cores representam os diferentes tipos morfológicos. *Painel inferior*: A fração de galáxias observadas pelo DR1 em relação a amostra-mãe. Retirado de Husemann et al. (2013), figura 2.



**Figura 2.2:** Este é o esquema com o *bundle* hexagonal com as 331 fibras de observação e mais 36 de amostra de céu. Retirado de Verheijen et al. (2004), figura 5.

vêm do PyCASSO que organiza as saídas da síntese de populações estelares resultantes da execução de cada espectro do IFU pelo STARLIGHT (Cid Fernandes et al. 2005) como descrito em Cid Fernandes et al. (2013) e também na próxima seção.

## 2.2 O *pipeline* PyCASSO

A beleza da espectroscopia de campo integral é poder unir imageamento e espectroscopia, obtendo-se assim a galáxia vista como um cubo de dados ( $x, y, \lambda$ ); para cada  $\lambda$  temos uma imagem, e para cada par de ascenção reta e declinação ( $x, y$ ) temos um espectro. !⊕j⊕! Pedir uma imagem chique de cubo para o André. Apesar de cada espectro poder ser analizado individualmente, nos cubos do CALIFA existe uma correlação entre pixels vizinhos devido a !⊕j⊕! seeing do céu e ao processo de observação. Como nas regiões mais afastadas do núcleo da galáxia o brilho superficial é menor, há uma queda na relação sinal-ruído (S/N) dos dados. De maneira que haja S/N constante, é feito um agrupamento de pixels (zonificação de Voronoi) nas regiões mais afetadas (ver (Cid Fernandes et al. 2013, sec. 3). Vale aqui frisar que a grande maioria das zonas comporta um pixel apenas. Para a galáxia NGC 2916, por exemplo, 93% das zonas possuem apenas 1 pixel e  $\sim 3\%$  com mais de 10 pixels agrupados, mesmo com a imposição de relação S/N em 20 (veja Tabela 2.1). Portanto o cubo original é transformado numa matriz de zonas e comprimentos de onda. Com esses dados é então possível realizar a síntese de populações estelares com o STARLIGHT, resolvendo espacialmente as propriedades físicas estelares das galáxias (Cid Fernandes et al. 2013, 2014; González Delgado et al. 2013).

### 2.2.1 E/S de dados no PyCASSO

O PyCASSO é uma biblioteca desenvolvida em Python para organizar os dados da síntese feita pelo STARLIGHT. A versão usada neste trabalho é a 0.9.3<sup>3</sup>. 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) de forma que a programação exploratória não precise se preocupar com as características de cada formato de armazenamento de dados. Por fim, existe uma camada construída para análise, com funções que retornam a indexação de cada zona para um par ( $x, y$ ), cálculos de

<sup>3</sup><http://minerva.ufsc.br/~andre/PyCASSO-0.9.3/>

**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           |

perfis radiais e azimutais, geometria, entre outras rotinas. Um programador facilmente pode adicionar mais rotinas como essas.

### 2.2.2 Exemplos de utilização

Ler um arquivo FITS é fácil com o PyCASSO, assim como o acesso aos dados. Na Figura 2.3 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.

Usando o `matplotlib`<sup>4</sup> podemos fazer gráficos facilmente acessando as matrizes e vetores do PyCASSO, como pode ser visto no programa na Figura 2.4 e sua imagem gerada (Figura 2.5). Os cálculos necessários para o PCA e também para a Tomografia PCA se tornam simples contas usando pacotes matemáticos de python.

Gerar perfis radiais ou axiais também são de fundamental importância para o estudo de diversas propriedades galáticas. Podemos ver na Figura 2.6 (e em sua imagem gerada 2.7) um exemplo de perfil radial executado através da função `radialProfile`.

Dentro de nossa colaboração já existem mais de dez pessoas utilizando a biblioteca Py-

<sup>4</sup><http://matplotlib.org>

```

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

```

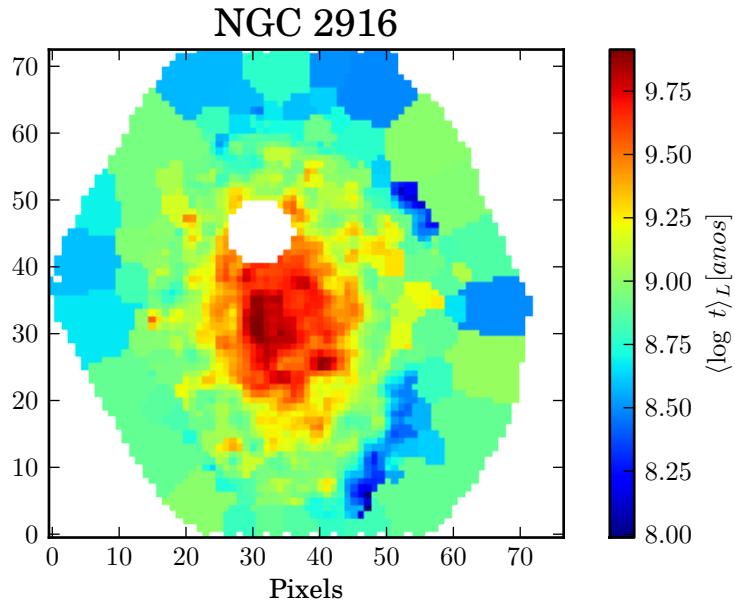
**Figura 2.3:** Exemplo de acesso aos cubo de dados por arquivo FITS e o cálculo da idade estelar média de uma galáxia.

```

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 2.4:** Programa para desenhar o mapa de idade estelar média ponderada pela luminosidade.



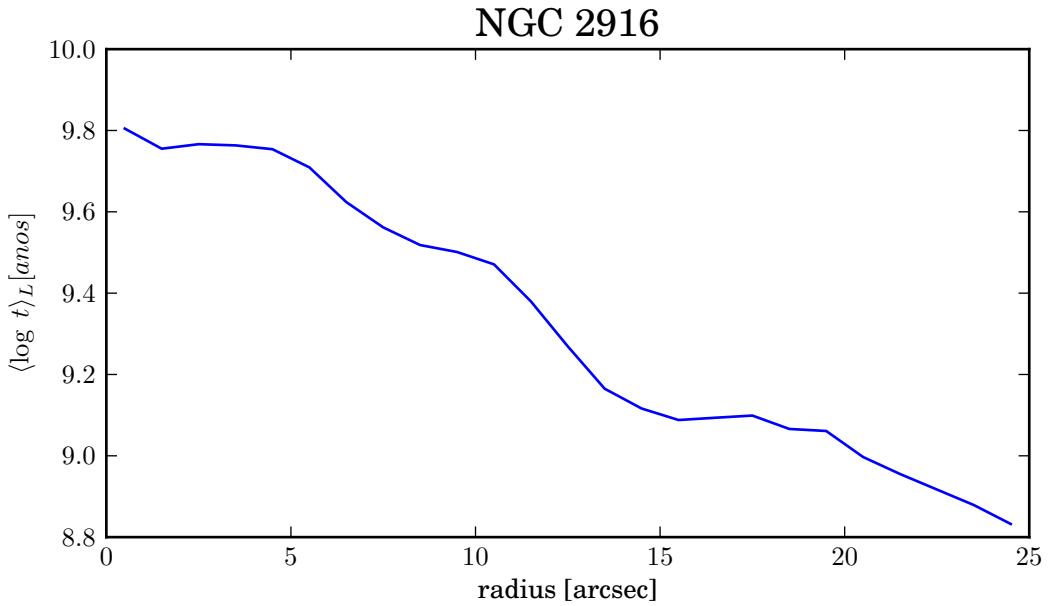
**Figura 2.5:** Mapa de idade estelar média ponderada pela luminosidade da galáxia NGC 2916 (CALIFA 277) gerado pelo programa da Figura 2.4.

```

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_imagem, bins, rad_scale = 1.0)
12
13 # Desenhar o perfil.
14 import matplotlib.pyplot as plt
15 plt.xlabel('radius [arcsec]')
16 plt.ylabel(r'\langle \log t \rangle_L [anos]')
17 cb = plt.colorbar()
18 plt.plot(bin_center, at_rad)

```

**Figura 2.6:** Programa para desenhar o perfil radial da idade estelar média ponderada pela luminosidade.



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

CASSO, com alguns artigos já publicados (Cid Fernandes et al. 2013, 2014; Pérez et al. 2013; González Delgado et al. 2013). Também há alguns usando indiretamente (Husemann et al. 2013; Iglesias-Páramo et al. 2013), uma tese de doutorado em curso e o presente trabalho.

# 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.

**!⊕j⊕!** Brincando com a própria técnica, o PCA em si é uma PC de um universo matemático-estatístico.

### 3.1 Principal Component Analisys

Baseada em encontrar os eixos com maiores variâncias em um conjunto de variáveis (no nosso caso, fluxos por lambda e por zona), a técnica PCA vem sendo de grande utilidade quando o assunto é estatística com muitas variáveis. Através de operações relativamente simples computacionalmente usando álgebra linear, é feito 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 e autovalores ( $\mathbf{e}_k$  e  $\lambda_k$ ) que maximizam o valor de  $\mathbf{e}_k^T \cdot \mathbf{C}_{cov} \cdot \mathbf{e}_k$  (ver Eq. 3.2) sujeito a restrição de que um autovetor deve ser ortogonal a qualquer outro da base ( $\mathbf{e}_i^T \mathbf{e}_j = 0$ ) e que todos devem ser normalizados ( $\mathbf{e}_i^T \mathbf{e}_i = 1$ ) (Jolliffe 2002, p. 5-6). No caso caso de PCA com espectros, cada um pode ser representado como um ponto num espaço  $\lambda$ -dimensional. Vários espectros formam uma nuvem de pontos. Assim, encontramos quais são os eixos mais significativos desse espaço em relação à variância, sujeitos às restrições acima. Fazemos esse cálculo encontrando

os autovetores e autovalores da matriz de correlação desse espaço (a matriz tem dimensão  $\lambda$ ) usando a biblioteca científica SciPy<sup>1</sup> (3.1), que encontra os autovetores e autovalores da matriz de correlação.

### 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. Os espectros observados estão armazenados em forma de uma matriz ( $n \times m$ ) com  $n$  zonas e  $m$  comprimentos de onda (`f_obs`<sup>2</sup> no PyCASSO).

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

Calculamos então o espectro médio de uma galáxia através da equação  $\langle \mathbf{F}_\lambda \rangle = (1/n) \sum_{i=0}^n f_{z_i\lambda}$  e então subtraímos a média de todos os espectros ( $\mathbf{I}_{z\lambda} = \mathbf{F}_{z\lambda} - \langle \mathbf{F}_\lambda \rangle$ ) para o cálculo da matriz de covariâncias usando um conjunto de dados com média zero. Vemos que a matriz de covariância possui dimensão  $\lambda \times \lambda$ .

$$\mathbf{C}_{cov} = \frac{[\mathbf{I}_{z\lambda}]^T \cdot \mathbf{I}_{z\lambda}}{n - 1} \quad (3.2)$$

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 PCs e os autovalores as variâncias. Isso feito, temos então o que necessitamos para iniciar o cálculo do Tomograma PCA. Muitas figuras de PCs e suas utilizações e diferenças nos pré-processamentos serão mostradas nos Capítulos ?? e 5, juntamente com a Tomografia PCA e as comparações com os parâmetros físicos da síntese de populações estelar.

---

<sup>1</sup><http://scipy.org/>

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

```
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 eigval e eigvect que são os autovalores e autovetores ordenados em forma decrescente.

## 3.2 Tomografia PCA

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

$$\mathbf{T}_{zk} = \mathbf{I}_{z\lambda} \cdot \mathbf{E}_{\lambda k} \quad (3.3)$$

Projetamos nossa matriz de observáveis com a média subtraída ( $\mathbf{I}_{z\lambda}$ ) na base das PCs ( $\mathbf{E}_{\lambda k}$ ). Na posse dessa nova matriz transformada e de um mapa que leve de zona para uma par de coordenada ( $z \rightarrow (x, y)$ ), podemos montar assim uma imagem. Cada imagem funciona como uma “fatia” de um cubo de dados expandido na nova base, assim formando a Tomografia PCA<sup>3</sup>, criada e assim batizada por Steiner et al. (2009), que em seu artigo faz um paralelo com fatias de um espaço tridimensional (tomograma do corpo humano, por exemplo) ou no espaço de velocidades (Tomografia Doppler). Cada “fatia” possui um autoespectro relacionado que, em conjunto, trazem novas perspectivas e ideias para a interpretação de ambos. A passagem de coordenadas  $z \rightarrow (x, y)$  está exemplificada na Figura 2.5 através da função de `zoneToYX` dentro do PyCASSO.

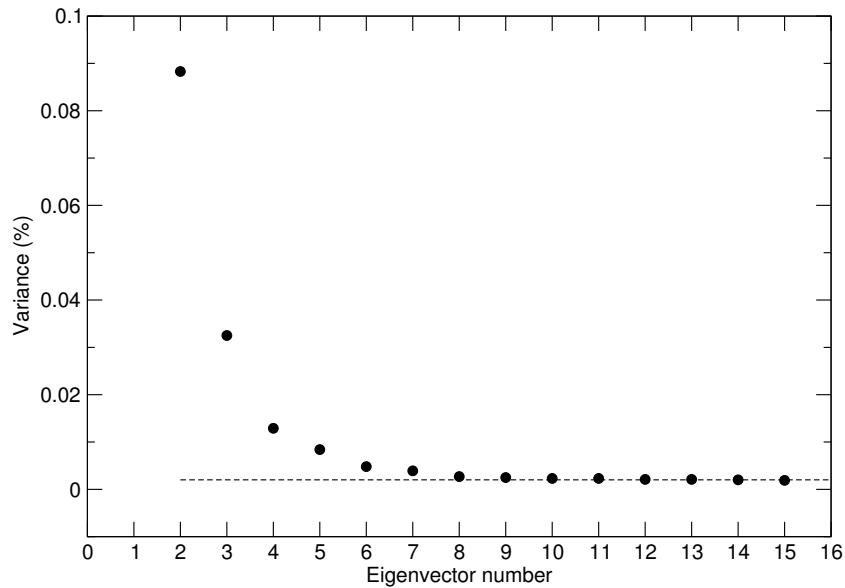
### 3.2.1 Evidências de linhas largas

No artigo citado anteriormente, através do estudo dos autoespectros, e suas respectivas imagens, da galáxia LINER (*Low Ionizations Nuclear Emission-line Region*) NGC 4736, foram encontradas evidências de *broad lines* nas regiões nucleares da galáxia. Quando temos uma fonte que é capaz de produzir linhas largas no espectro é sinal da existência de um SMBH (*Super Massive Black Hole*). Cid Fernandes et al. (2004) mostra que a subtração bem detalhada das populações estelares nos espectros ajuda a encontrar linhas largas mais fracas em Seyferts-II, que são aquelas que possuem linhas estreitas, ajudando assim em na classificação desses objetos como tipo I ou II. O PCA, juntamente com a Tomografia PCA, fazem esse papel da subtração das populações estelares sem haver nenhuma parametrização.

Foram estudados os primeiros 8 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 com E1) possui 99.74% da variância e reproduz comportamento do gás e da população estelar somados (Figura 3.3). O segundo contribui com

---

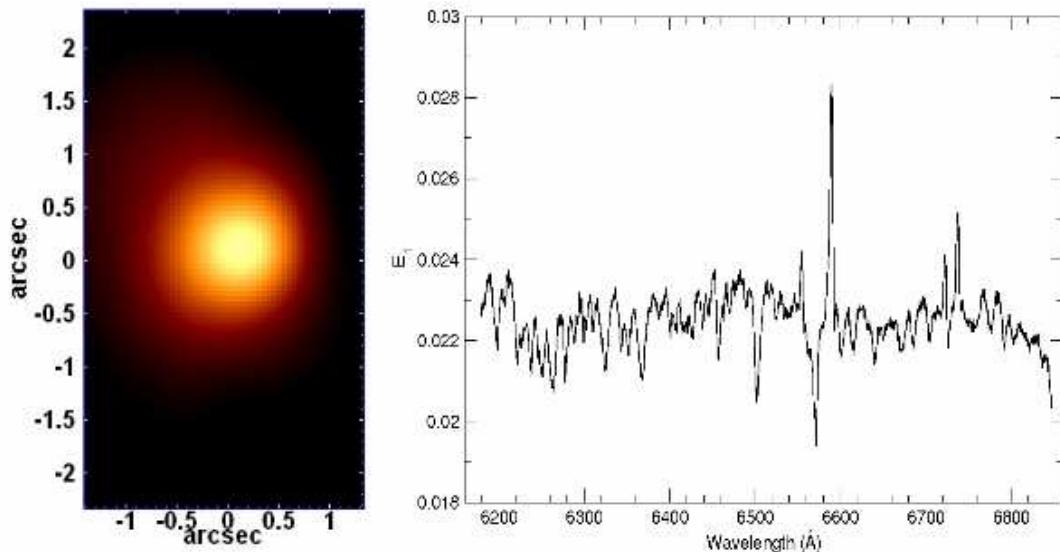
<sup>3</sup><http://www.astro.iag.usp.br/~pcatomography/>



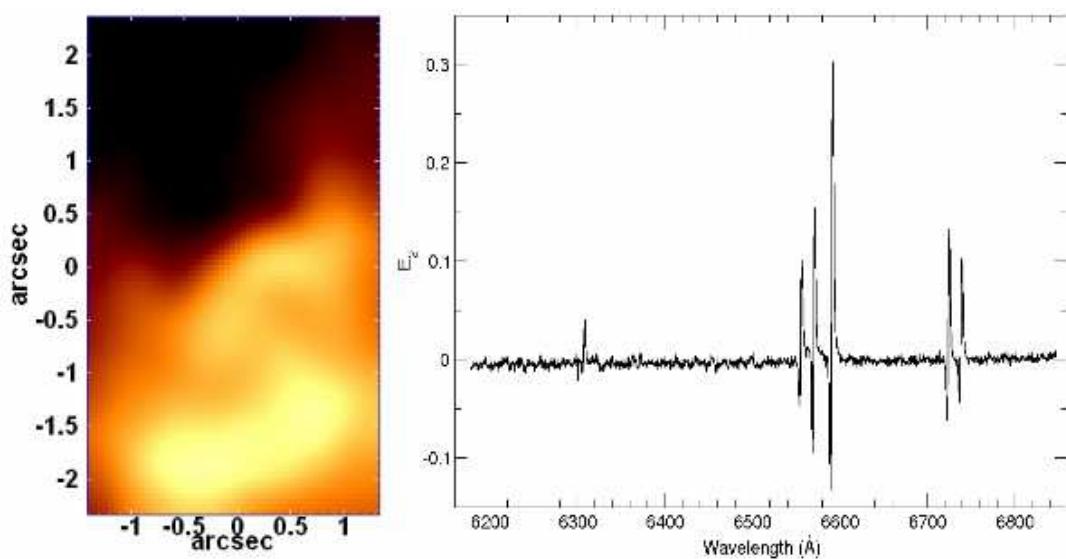
**Figura 3.2:** Scree test das primeiras 16 PCs do cubo de espectros da região central da galáxia NGC 4736. Retirado de Steiner et al. (2009, fig. 1).

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). Mas é no terceiro (0.032% da variância) que mostra evidências de uma emissão larga de H $\alpha$  (Figura 3.5). Essa assinatura é uma evidência típica de objetos com um AGN associado a um SMBH (Seyfert-I).

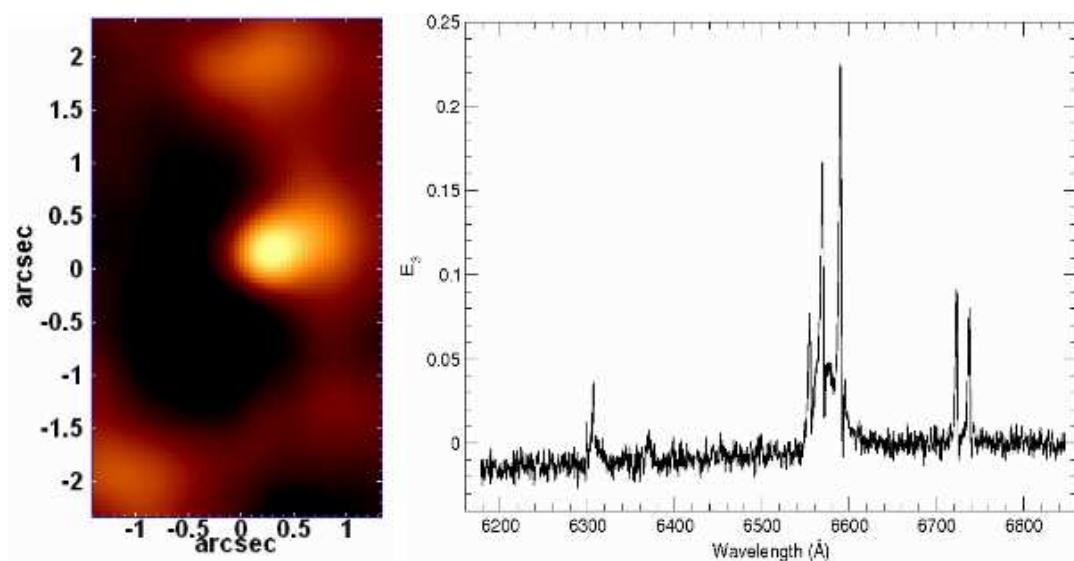
Com isso em mãos podemos estudar a Tomografia PCA nas galáxias do CALIFA. Temos disponíveis várias galáxias com os cubos de dados COMBO e aplicaremos essa técnica às galáxias presentes na Tabela 2.1 no Capítulo 5.



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



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



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

# Capítulo 4

## PCA em IFS - (different flavors?!? FIXME PLEASE)

**[FIXME]** Com o cubo de espectros já preparados pelo QBICK e com as limitações em comprimento de onda à mão, estamos prontos para executar a Tomografia PCA. Como o PCA é uma técnica em qual calcula-se os eixos que, através da variância, melhor expandem sua base de dados, é natural que qualquer alteração na variância entre os espectros, resultará 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 ao ponto de alguns pré-processamentos serem de extrema necessidade para interpretação dos resultados. Por fim faremos uma espécie de engenharia reversa, que nos auxilia na busca do sentido físico das principais componentes resultantes. Neste Capítulo todos as análises serão feitos usando a galáxia NGC 2916 ( $z \sim 0.01244$  Hubble Type - Sbc). Mais dados sobre elas as outras galáxias usadas nesse trabalho virão no Capítulo 5.

### 4.1 Normalização

Como exemplo inicial apresentamos as cinco primeiras PCs e seus respectivos tomogramas do cubo de espectros observados da galáxia NGC 2916 (Figura 4.1). O tomograma é formado pelo “peso” da PC em cada zona. A primeira PC é praticamente o espectro médio, que fica evidente comparando com o espectro médio em cinza no segundo eixo vertical (à direita).

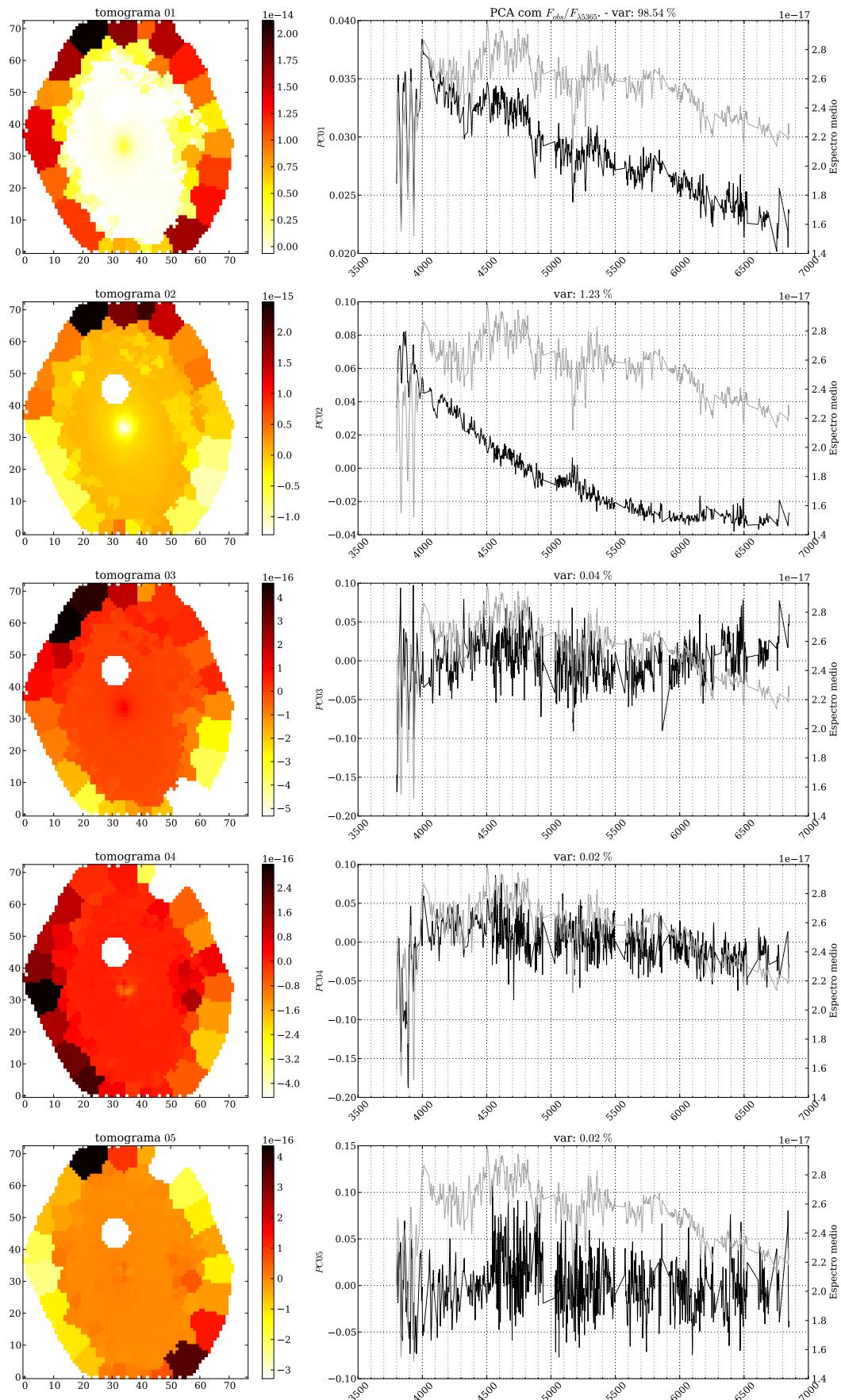
O segundo autoespectro se assemelha com um espectro de população muito jovem. Com o auxilio de seu tomograma vemos que para as zonas mais internas da galáxia (onde se presume que as populações sejam mais velhas [citation needed]) o seu peso é negativo. De forma contrária, em regiões mais afastadas do núcleo, sobre os braços espirais, 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 e no próximo vamos falar mais sobre como melhor interpretar as PCs fisicamente.

Da mesma forma, fazemos o PCA para o mesmo cubo, mas com os espectros normalizados da forma que comentamos em [FIXME] (Capítulo 2) (Figura 4.2). À primeira vista os resultados podem ser parecidos com o anterior deslocando uma componente acima, ou seja, o mesmo PCA mas sem a primeira PC. Mas essa é só uma primeira impressão. As propriedades físicas de uma galáxia geralmente possuem simetria radial ou axial, sendo assim extremamente correlacionadas no espaço. Matematicamente, a maior variância nos espectros espacialmente é sua amplitude, aparecendo na forma de componente principal como visto na análise feita sem normalização. Essa componente, 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 informação alguma a análise das populações, funciona como um fator de escala apenas, afetando as componentes seguintes. Esses vestígios de amplitude aparecem de forma mais evidente quando comparamos a PC3 no caso sem normalização com a PC2 no caso com normalização. Podemos ver um padrão de rotação na PC2 da Figura 4.2. Com uma saturação na escala de cores no tomograma da PC3 da Figura 4.1 diminuímos o efeito do fator de escala ainda presente nessa componente, de modo que ficasse mais evidente o padrão de rotação. O espectro médio já é a informação necessária que precisaríamos para entender esse fator de escala nas amplitudes dos espectros, portanto a primeira componente para o caso sem normalização não trás informação adicional. Veremos a seguir que seu tomograma também não trás maiores informações.

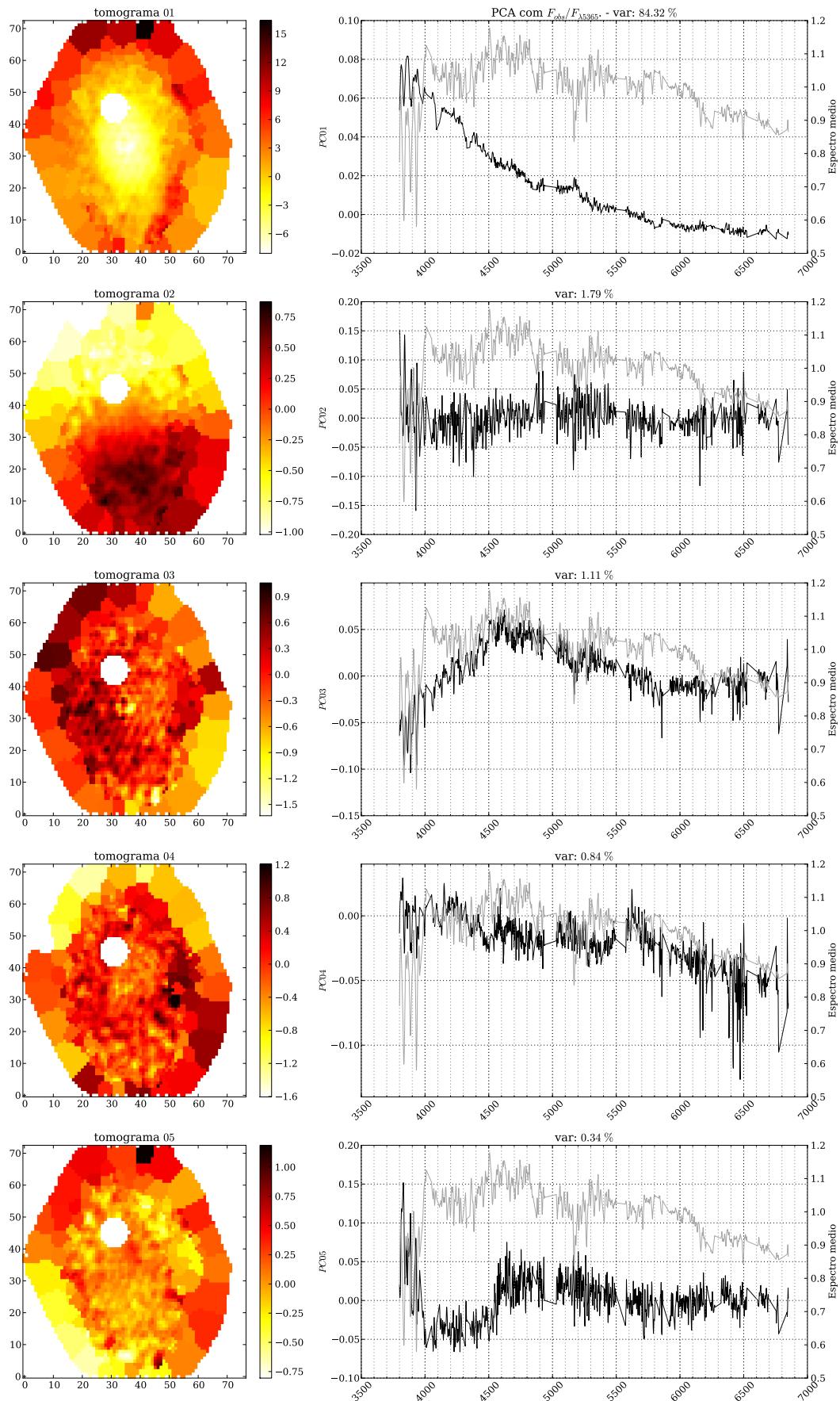
Imagine uma galáxia composta inteiramente pela mesma população estelar, em repouso, distribuídas da mesma maneira no espaço. Ou seja, em qualquer ponto da galáxia o espectro é o mesmo. Um PCA nessa galáxia hipotética nos mostraria apenas uma componente relevante. Adicione então a essa galáxia uma função de densidade de massa em função do raio, permitindo que de uma posição para outra a quantidade dessa determinada população se altera (mudando o brilho superficial da região). Uma componente nova irá surgir na sua análise

PCA, mostrando que existe uma variância agora numa componente de escala (amplitude) nos espectros. Mas o que essa componente de escala nos diz sobre a física da população estelar existente? Essa componente seria um desperdício de variância para uma análise de populações estelares de uma galáxia.

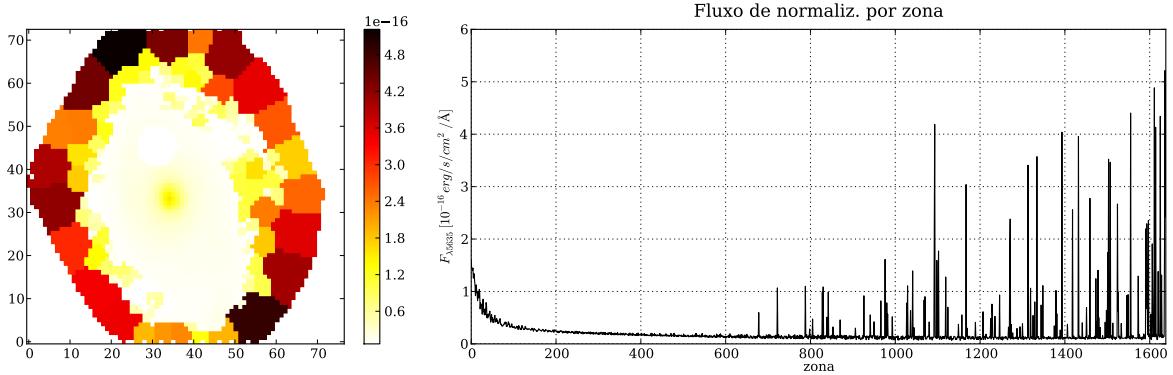
No caso do CALIFA, com um *FoV* abrangendo praticamente toda a galáxia, esse efeito de amplitude se acentua muito devido ao brilho superficial mais intenso nas zonas centrais da galáxia em comparação com as mais afastadas, assim adicionando uma grande variância descartável entre as zonas. Descartável pois não trazem informação nova para a nossa análise. Essas diferenças em amplitude não nos dizem nada sobre as populações estelares. Comparemos agora a primeira PC do caso sem normalização (Figuras 4.1) e a imagem mais à esquerda na Figura 4.3 formada pelos fluxos para normalização por zona. Podemos notar que o primeiro autoespectro (e seu respectivo tomograma) no caso sem normalização mostra exatamente esse fator de escala. Seu tomograma mostra que ela é claramente um fator de escala (que pode ser considerado um fator de brilho, ou de amplitude, nos espectros). Por esse motivo, em nossas análises daqui para frente, usaremos o cubo com os espectros normalizados.



**Figura 4.1:** Cinco primeiras PCs (e seus respectivos tomogramas) resultantes da Tomografia PCA aplicado aos espectros sem normalização da galáxia NGC 2916.



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



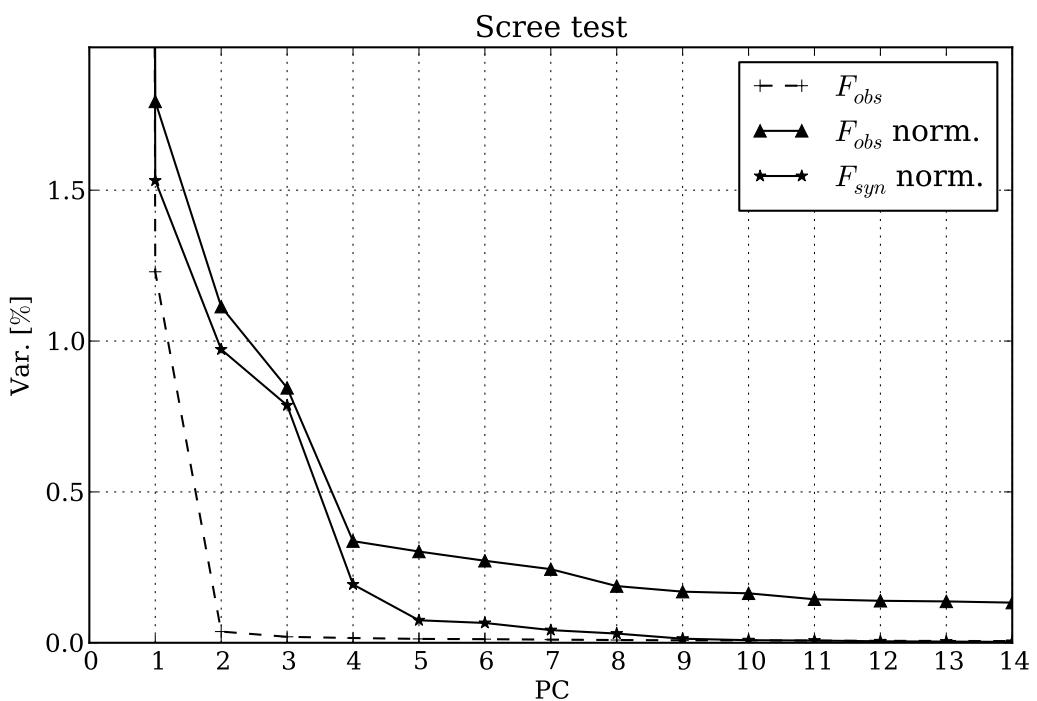
**Figura 4.3:** A imagem à esquerda e a superior direita, mostram o fluxo usado para a normalização de cada espectro em cada pixel (à esquerda) ou zona (superior direita). Na imagem inferior direita temos um histograma para valores do fluxo de normalização.

## 4.2 Fluxos observados e sintéticos

Com o resultado da síntese de populações estelares já organizado para as galáxias do CALIFA, realizamos o PCA no cubo de espectros observados e no de espectros sintéticos. A grande diferença é que nos espectros sintéticos estão contidas apenas as informações sobre populações estelares<sup>1</sup>. Como os espectros sintéticos não possuem as assinaturas dos equipamentos observacionais, dos processos para subtração do céu, ruídos e afins, quando analisadas pela técnica de PCA vemos que as informações se condensam em menos PCs. Comparando os dois *scree tests* na Figura 4.4 vemos que para o caso com os espectros sintéticos a curva converge mais rápido ao zero de variância, mostrando que temos as informações mais compactadas nas primeiras PCs quando comparadas ao caso com os espectros observados. Observando o caso sem normalização, plotado no gráfico com linha pontilhada, vemos que o efeito causado pelo fator de escala (PC1) diminui a “importância” (i.e., variância) das demais PCs. As cinco primeiras PCs e seus tomogramas provenientes do cubo com os fluxos sintéticos normalizados da galáxia NGC 2916 aparecem na Figura 4.5.

Comparando os dois vemos que os resultados para os espectros sintéticos parecem ser mais “limpos”, pois não existem ruídos. E são! Como são espectros gerados a partir de uma base teórica para diferentes idades e metalicidades de populações estelares, também não possuem nenhuma assinatura instrumental. Por esse fato acabamos descobrindo uma assinatura pre-

<sup>1</sup>Suavização, correções por poeira e cinemática também são feitas nos espectros no processo de síntese. Mais detalhes em Cid Fernandes et al. (2005)



**Figura 4.4:** Scree test para 3 análises PCA da galáxia NGC 2916 (CALIFA 277). Com marcações de triângulos vemos as PCs resultantes do PCA com os espectros observados normalizados. As variâncias das PCs marcadas com estrela representam o PCA com os espectros sintéticos normalizados. Para comparação plotamos as PCs do caso sem normalização usando linha pontilhada.

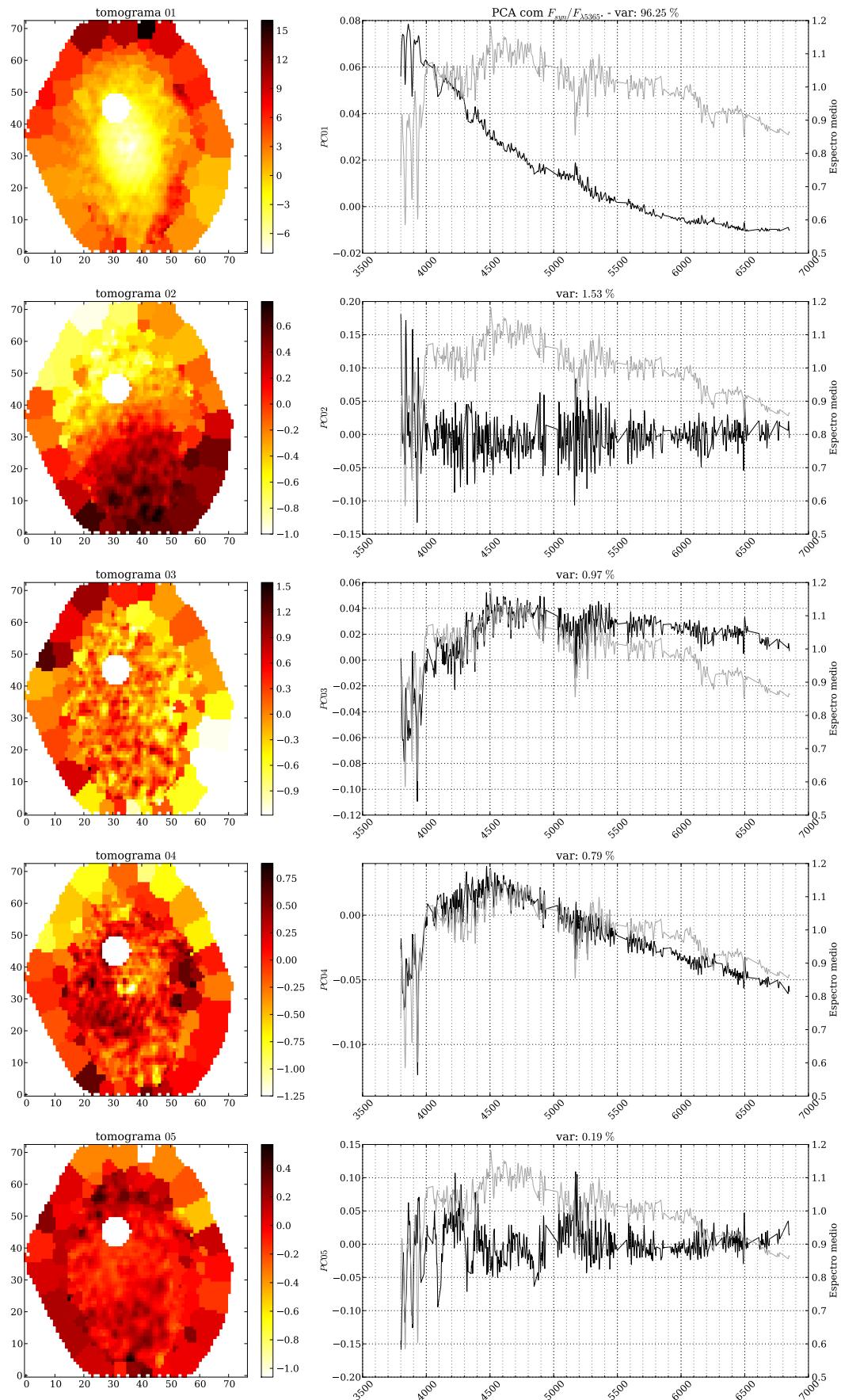
sente em quase todas as componentes geradas pelo PCA usando os dados observados. Como comentamos no Capítulo 2.1 utilizamos o cubo de dados COMBO, gerado a partir da união do V500 com o V1200. Como os espectros V1200 possuem espectros com maior resolução do que os do V500 (V1200 - FWHM  $\sim$  2.3 Å; V500 - FWHM  $\sim$  6 Å), 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 Å). A PC5 da Figura 4.2 mostra um degrau grande entre 4500 e 4600 Å qual parece mostrar essa diferença de comportamento entre as duas versões originais antes da formação do COMBO. Nessa componente fica mais evidente, mas traçando uma linha imaginária verticalmente por esse comprimento de onda fica mais fácil de perceber uma diferença de comportamento nos dados. Já nas componentes sintéticas<sup>2</sup> não se vê essa mudança de comportamento no autoespectro.

A cinemática também adiciona uma variância descartável aos dados. 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  $\lambda$ . Esses efeitos cinemáticos não estão nos trazendo informação alguma para o estudo das populações estelares. Causam um grande disperdício de variâncias, sempre aparecendo nas primeiras PCs. Uma manipulação através do vetor de população<sup>3</sup> criado pela síntese, juntamente com os espectros da base, pode nos ajudar a criar os espectros sintéticos sem nenhuma correção por cinemática e poeira de maneira que se execute o PCA sem o disperdício de variância dessas componentes cinemáticas. Também é importante lembrar que existem métodos mais eficazes e direcionados para a determinação de tais propriedades cinemáticas. Nas galáxias presentes no CALIFA não poderia ser diferente, portanto os espectros aparecem com linhas deslocadas para o azul (*blue-shifted*) ou para o vermelho (*red-shifted*) dependendo da velocidade de rotação projetada. A dispersão de velocidades em cada ponto da galáxia também pode causar alargamento ou estreitamento das linhas.

---

<sup>2</sup>Componentes geradas pelo PCA nos cubos de espectros sintéticos.

<sup>3</sup>O vetor de população diz o quanto de cada tipo de estrela entra na receita para construir o espectro sintético.



**Figura 4.5:** Cinco primeiras PCs (e seus respectivos tomogramas) resultantes da Tomografia PCA aplicado aos espectros sintéticos normalizados da galáxia NGC 2916.

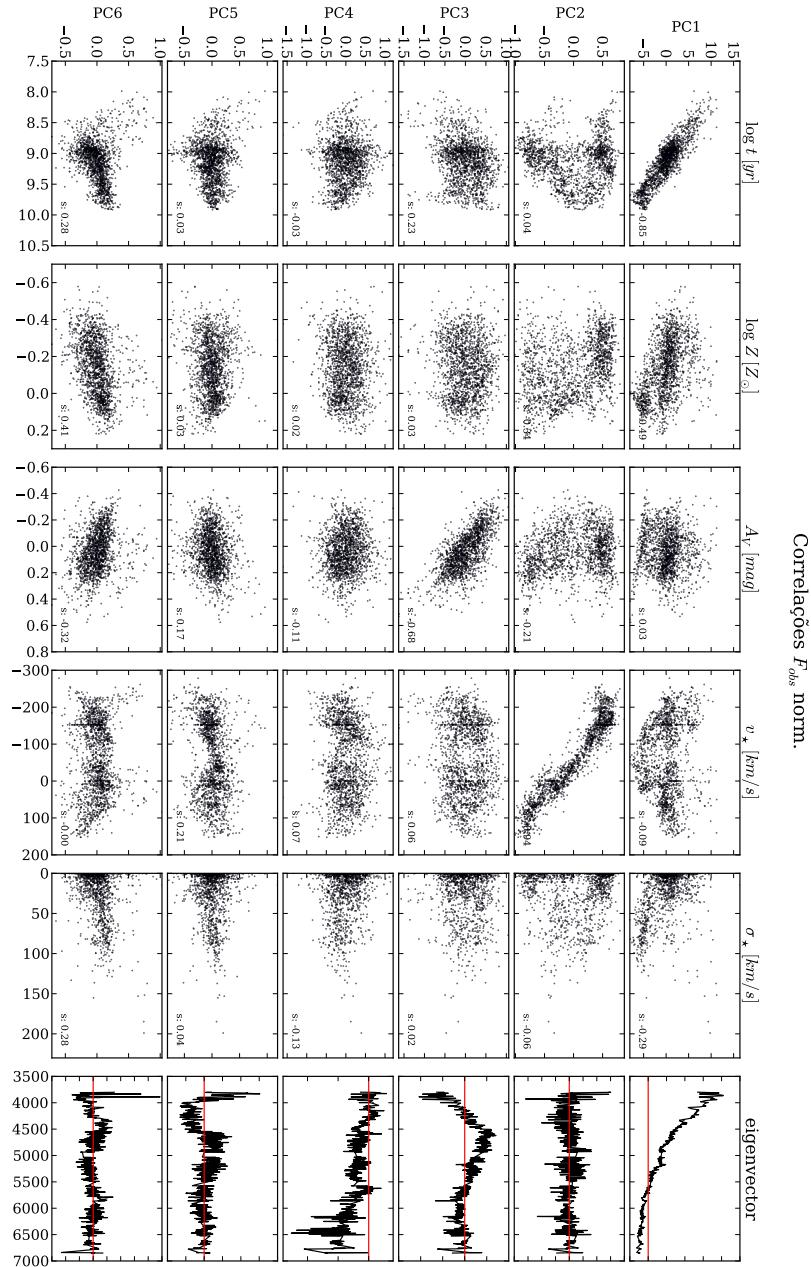
## 4.3 Comparando as PCs com o *STARLIGHT*: engenharia reversa

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 pelo PCA. Nas Figuras 4.6 e 4.7 vemos as correlações entre alguns parâmetros físicos ( $\log t$ ,  $\log Z/Z_{\odot}$ ,  $A_V$ ,  $v_{\star}$ ,  $\sigma_{\star}$ ) e o peso de cada PC por zona (tomograma versus parâmetro físico) para a galáxia NGC 2916. O número presente em cada gráfico é o coeficiente de correlação por *rank* de Spearman. Escollhamos o coeficiente de Spearman pois este é aparamétrico e, ao contrário do coeficiente de Pearson, não pressupõe nenhuma relação linear.

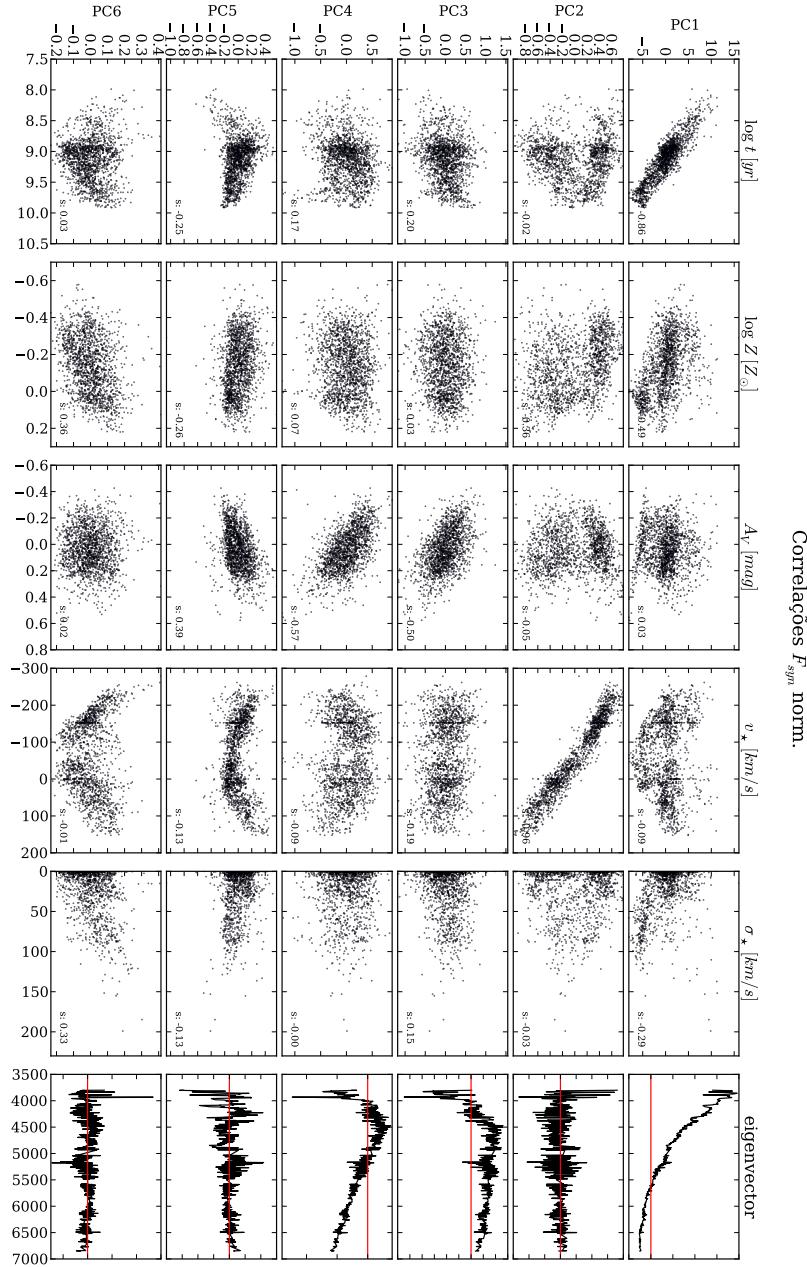
Em cada coluna temos uma quantidade física e na última coluna temos o autovetor representado por aqueles pontos. tanto para os dados observados quanto sintéticos vemos que a primeira PC representa basicamente um fator de idade. Possui correlação também com a metalicidade. A segunda PC também em ambos os casos representa uma forte correlação com a velocidade estelar. 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 spearmann baixo.

Outra forma de se observar essas correlações é graficando os dados nos eixos das PCs, colorindo os pontos por determinado parâmetro físico. No Apêndice estarão todos os gráficos com todos os parâmetros físicos. Como exemplo aqui deixamos as Figuras 4.8 e 4.9 que mostram as PCs para o caso observado e sintético, coloridos pelo  $A_V$ .

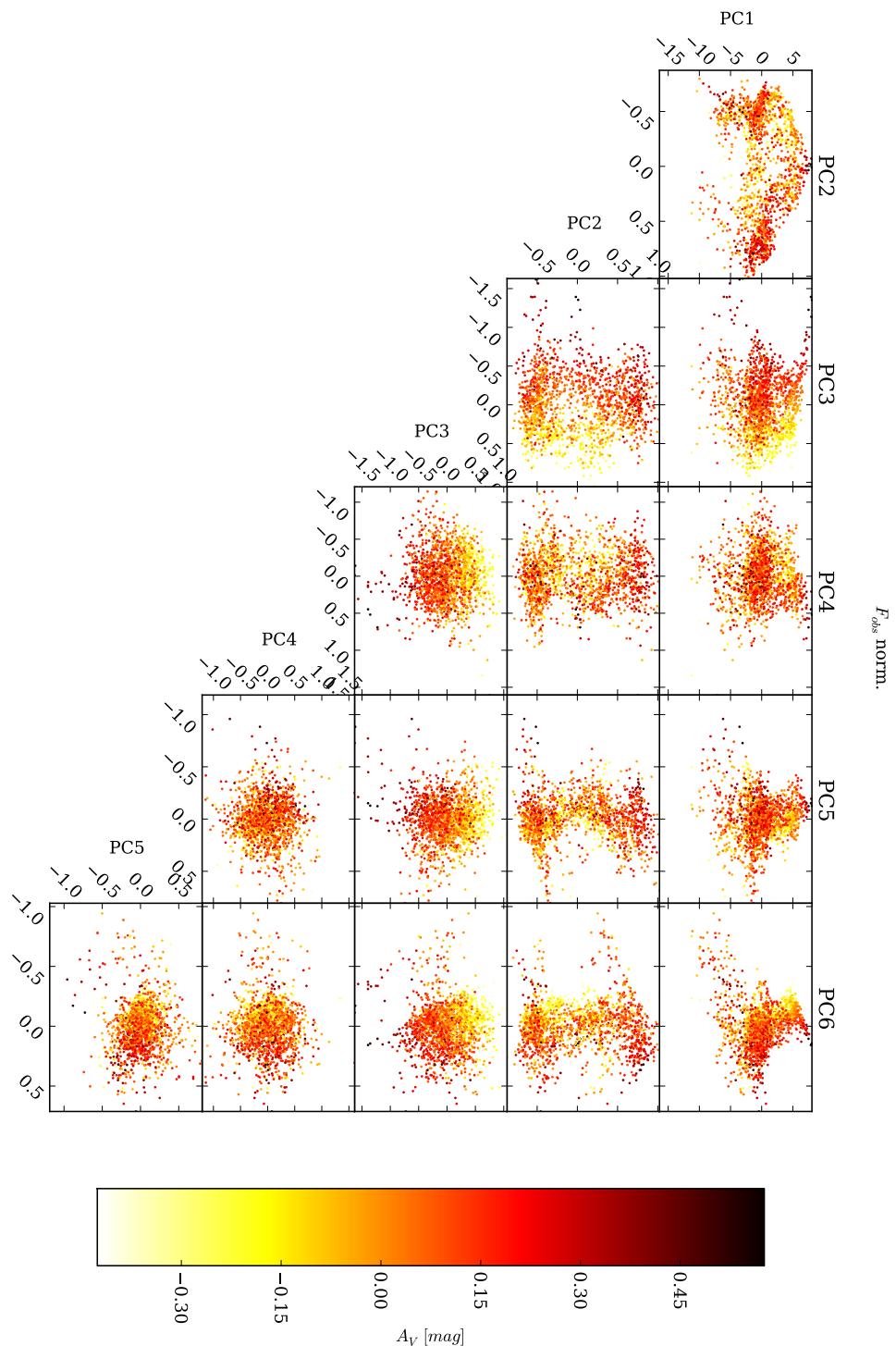
A Figura 4.10 mostra também o efeito que a falta de normalização faz com o resultado do PCA. Esse fator de amplitude que não é removido pela normalização se concentra principalmente na primeira PC, mas ainda deixa vestígios se misturando com as outras PCs fazendo com que haja correlação entre cada PCs e quase todos os parâmetros físicos.



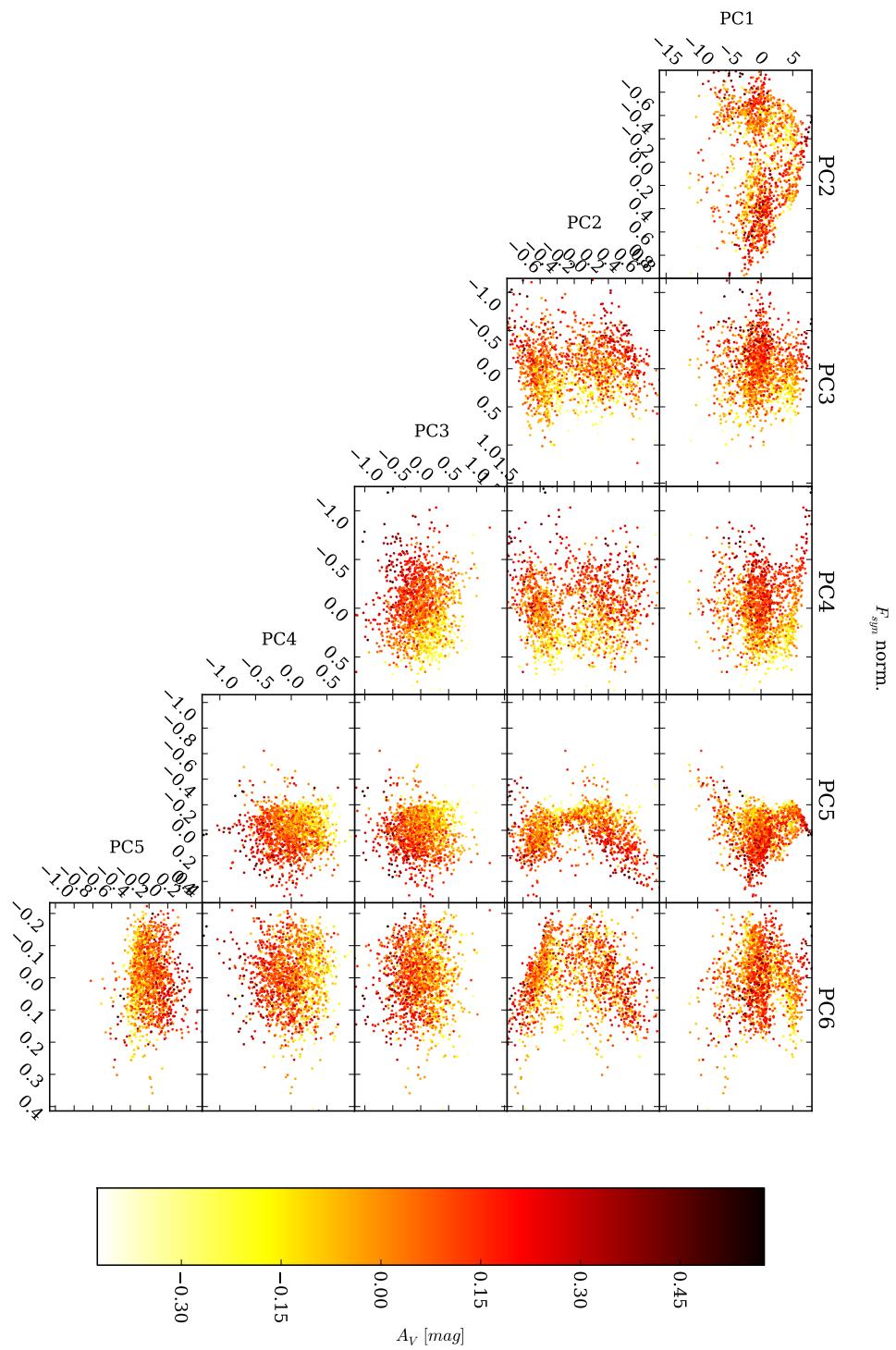
**Figura 4.6:** Correlações entre os pesos por zona das seis primeiras PCs do PCA feito para o cubo com os dados observados normalizados e cinco parâmetros físicos. Pela ordem de colunas da esquerda para direita temos  $\log t$ ,  $\log Z/Z_{\odot}$ ,  $A_V$ ,  $v_{\star}$ ,  $\sigma_{\star}$ . Na última 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.



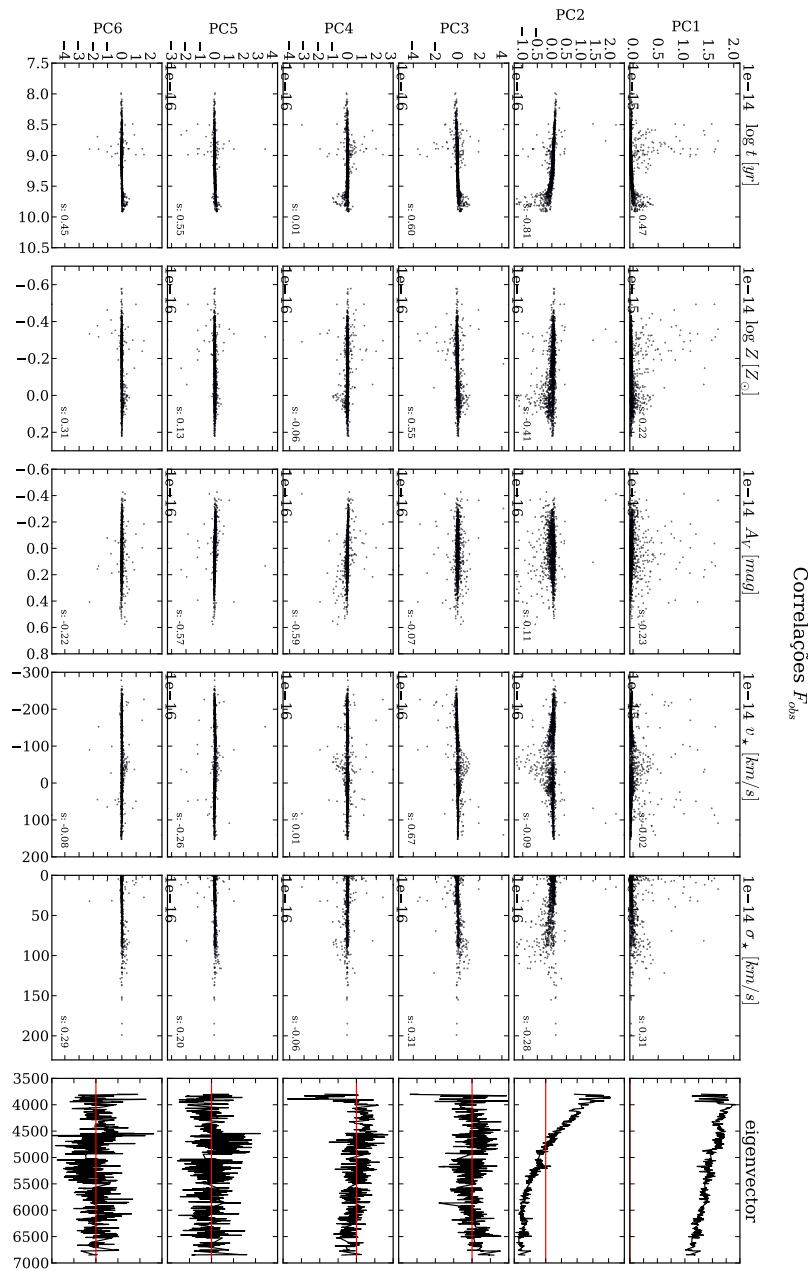
**Figura 4.7:** Correlações entre os pesos por zona das seis primeiras PCs do PCA feito para o cubo com os dados sintéticos normalizados e cinco parâmetros físicos. Pela ordem de colunas da esquerda para direita temos  $\log t$ ,  $\log Z/Z_{\odot}$ ,  $A_V$ ,  $v_{\star}$ ,  $\sigma_{\star}$ . Na última 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 4.8:** Dados das zonas no espaço das PCs (com espectros observados) coloridos pela extinção ( $A_V$ ).



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



**Figura 4.10:** Correlações entre os pesos por zona das seis primeiras PCs do PCA feito para o cubo com os dados observados e cinco parâmetros físicos. Pela ordem de colunas da esquerda para direita temos  $\log t$ ,  $\log Z/Z_{\odot}$ ,  $A_V$ ,  $v_{\star}$ ,  $\sigma_{\star}$ . Na última coluna temos o autoespectro para ajudar na visualização. A linha em vermelho no gráfico do autoespectro serve para identificar o zero.

.....

Caso as grandezas físicas analisadas em uma galáxia fossem não-correlacionadas seria muito mais fácil fazer uma comparação entre PCs e grandezas físicas, mas geralmente não é o que ocorre. Para colhermos informações sobre os objetos no céu só temos duas formas: através imagens ou de espectros. Em ambas formas, diferentes efeitos físicos causam efeitos semelhantes nas cores (imagem) ou nos espectros. Esses efeitos fazem com que alguns dos parâmetros físicos fiquem extremamente correlacionados. **!OJO!** [citation needed] Gostaria de adicionar aqui algumas referências como alguns dos papers do SEAGal/Starlight que fala sobre as degenerescencias de idade e metalicidade e algumas outras grandezas correlacionadas. Coloquei essa parte aqui pensando naquela crítica que o cara me fez lá no México sobre as grandezas em astrofísica serem extremamente correlacionadas, mas não sei até que ponto é legal comentar sobre isso visto que estamos analisando as coisas via correlação... talvez essa discussão possa ficar mais pro Cap 6 nas conclusões..

.....

**[FIXME]** **DAQUI PRA BAIXO VAI PRO CAP 2** Antes dos espectros chegarem ao PyCASSO, todas as informações de *flags*<sup>4</sup> em *bad pixels* e linhas telúricas<sup>5</sup> são criadas em um *pipeline* de pré-processamento chamado QBICK. Esse *pipeline* também prepara os cubos para a execução do STARLIGHT, que posteriormente serão organizados pelo PyCASSO, definindo as zonas de Voronoi, a reamostragem em  $\lambda$  e colocando os espectros em repouso usando o *redshift* calculado dentro dos 5" centrais da galáxia. Todas essas informações e pré-processamentos provenientes do QBICK são herdadas pelo PyCASSO e já estão contidas em seus cubos de espectros.

Primeiro todos os espectros são limitados ao intervalo de 3800 a 6850 Å. Após essa limitação, fazemos uma estatística com todos os *bad pixels* e linhas telúricas 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. Podemos ver o efeito desses pré-processamentos nos espectros através dos primeiro e segundo espectros na Figura 4.11.

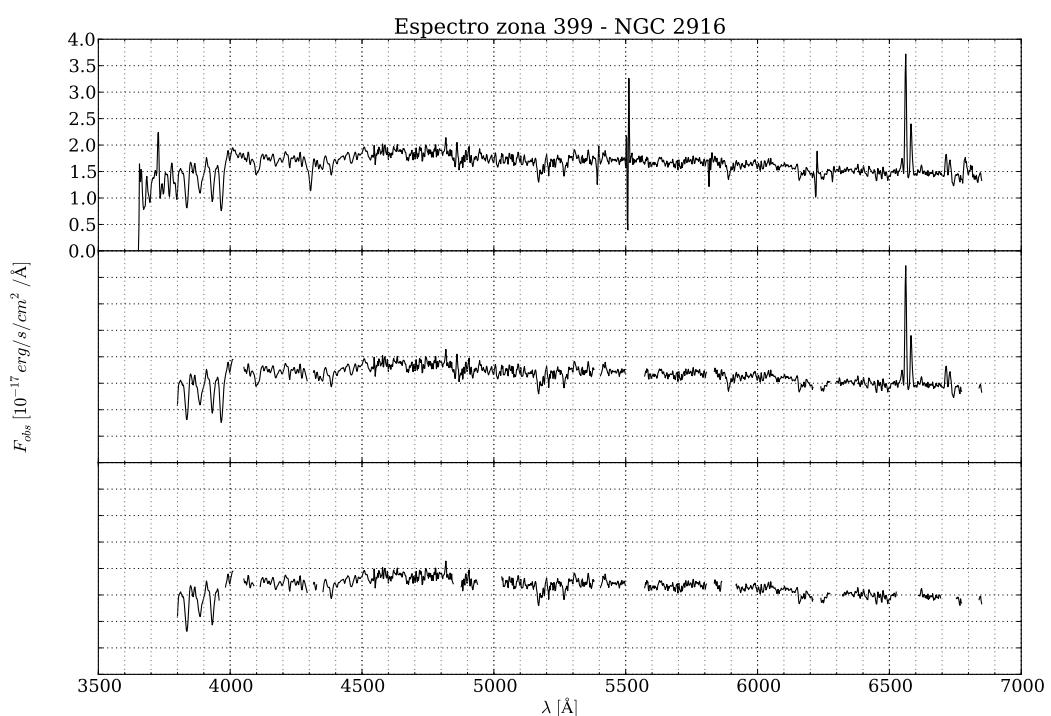
.....

Nos espectros, além dos *bad pixels* e linhas telúricas, podemos mascarar regiões desnecessárias para determinada investigação científica. Nosso foco é o estudo das populações

---

<sup>4</sup>Marcações.

<sup>5</sup>Linhas de emissão ou absorção referentes à atmosfera.



**Figura 4.11:** Espectro da zona 399 da galáxia NGC 2916 (CALIFA 277). Acima está o espectro completo. No segundo vemos o espectro com linhas telúricas e bad-pixels removidos, além do limite de intervalo em comprimento de onda de 3800 a 6850 Å. No espectro mais abaixo, além das partes removidas no segundo, estão removidas também as mesmas linhas de emissão mascaradas na síntese de populações estelares.

estelares, portanto necessitamos que as linhas de emissão, geralmente associadas ao gás presente nas galáxias **[FIXME]** sejam removidas do espectro. Para que possamos fazer correlações entre os resultados do PCA e as propriedades físicas obtidas pela síntese através do *STARLIGHT*, em nossos espectros também são mascaradas todas as regiões do espectro que não entram na síntese ( $H\epsilon$ : de 3960 a 3980 Å;  $H\delta$ : de 4092 a 4112 Å;  $H\gamma$ : de 4330 a 4350 Å;  $H\beta$ : de 4848 a 4874 Å; [O III]: de 4940 a 5028 Å; He I e NaD: de 5866 a 5916 Å;  $H\alpha$  e [N II]: de 6528 a 6608 Å; S II: de 6696 a 6752 Å). O espectro após todas máscaras pode ser visto no terceiro espectro (o mais abaixo) da Figura 4.11.

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

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

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

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

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

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

# **Capítulo 5**

## **Resultados**

**!obj!** Aplicando PCA e Tomografia PCA para 4 galáxias espirais, 2 elípticas e 2 mergers.

# **Capítulo 6**

## **Conclusões e perspectivas**

### **6.1 Este trabalho**

**!○j○!**

### **6.2 Trabalhos futuros**

#### **6.2.1 Linhas de emissão e intervalos específicos em comprimento de onda**

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

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

# Apêndice A

## [FIXME] Python

!○j○!

# Referências Bibliográficas

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

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

- Skrutskie, M. F., Cutri, R. M., Stiening, R., Weinberg, M. D., Schneider, S., Carpenter, J. M., Beichman, C., Capps, R. et al. 2006, AJ, 131, 1163
- Steiner, J. E., Menezes, R. B., Ricci, T. V., & Oliveira, A. S. 2009, MNRAS, 395, 64
- Verheijen, M. A. W., Bershady, M. A., Andersen, D. R., Swaters, R. A., Westfall, K., Kelz, A., & Roth, M. M. 2004, Astronomische Nachrichten, 325, 151
- York, D. G., Adelman, J., Anderson, Jr., J. E., Anderson, S. F., Annis, J., Bahcall, N. A., Bakken, J. A., Barkhouse, R. et al. 2000, AJ, 120, 1579