

Universidade de São Paulo
Instituto de Astronomia, Geofísica e Ciências Atmosféricas
Departamento de Astronomia

Lilianne Mariko Izuti Nakazono

Separação de estrelas e quasares em imagens multiespectrais

São Paulo

2016

Lilianne Mariko Izuti Nakazono

Separação de estrelas e quasares em imagens multiespectrais

Monografia apresentada ao Departamento de
Astronomia do Instituto de Astronomia, Geofísica
e Ciências Atmosféricas da Universidade de
São Paulo como requisito para conclusão de
curso.

Área de Concentração: Astronomia

Orientador(a): Prof.^(a) Dr.^(a) Claudia Lucia
Mendes de Oliveira e Prof.^(a) Dr.^(a) Nina Su-
miko Tomita Hirata

São Paulo

2016

Para Dirce e Hanako.

Agradecimentos

Primeiramente, agradeço à minha família pelo apoio, carinho e compreensão nesta jornada; às minhas orientadoras Claudia e Nina por toda a disposição e ensinamentos; ao professor João E. Steiner e ao Alberto Molino que colaboraram significativamente neste projeto; à FAPESP, pelo apoio financeiro, sob o projeto nº: 2014/10566-4; à todos que me sugeriram melhorias neste projeto; aos avaliadores e ao Instituto de Astronomia, Geofísica e Ciências Atmosféricas pelo prestígio com o prêmio de destaque no SIICUSP por este projeto; à todos os professores e professoras da Universidade de São Paulo e do ensino básico ao médio por terem sido essenciais na minha formação; aos funcionários do IAG e, em especial, do departamento de Astronomia que sempre foram cordiais e prestativos; por fim, porém não menos importante, aos grandes amigos que fiz nestes anos da graduação e que fizeram esse árduo caminho se tornar mais suave, em especial à Stela.

*“Seja como os pássaros que, ao pousarem um instante sobre ramos muito leves,
sentem-nos ceder, mas cantam! Eles sabem que possuem asas.”*

Victor Hugo

“Se não encontras a alegria nesta terra, procura-a irmão para além das estrelas.”

Platão

Resumo

Com o intuito de encontrar novos candidatos a quasar, utilizamos a Análise de Componentes Principais (ACP) em uma amostra de quasares conhecidos e de objetos *a priori* considerados como estrelas. Estes grupos foram amostrados do ALHAMBRA *Survey*, que possui observações em 20 filtros no óptico e 3 em infravermelho próximo. A partir dessas observações determinamos parâmetros que tenham comportamentos diferentes entre as duas classes (por exemplo, H_α e H_β) a serem utilizadas como *input* da ACP. Pelos gráficos entre componentes principais distinguimos regiões de concentração de estrelas e de quasares e, após um refinamento da análise, pudemos determinar 985 candidatos a quasar. Após esta etapa, elaboramos e testamos um método de estimação de *redshift* a fim de aplicar nestes candidatos e verificar a classificação dos mesmos.

Abstract

Aiming to find new quasars candidates, we applied the Principal Component Analysis (ACP) on samples of known quasars and objects first considered as stars. These groups were sampled from ALHAMBRA *survey* which observed the sky with 20 optical filters and 3 near-infrared filters. We selected parameters that distinguish both classes (*e.g.* H_α and H_β) and they were our input to ACP. Plots between principal components allowed us to view different locus for stars and quasars. After some refinement, we identified 985 quasars candidates. Afterwards, we elaborated and tested a method to estimate redshifts in order to classify those candidates.

Lista de Figuras

1.1	Imagem do QSO PHL 909	19
1.2	Conjunto de filtros do ALHAMBRA <i>survey</i>	20
1.3	Conjunto de filtros do J-PAS	21
2.1	Exemplo de dois quasares encontrados nos dados do ALHAMBRA Survey (DS9)	24
3.1	Matriz de autocorrelação entre os dez parâmetros utilizadas na ACP	26
3.2	Gráfico de dispersão entre componentes principais no caso do modelo linear	27
3.3	Gráfico de dispersão entre CP3 e CP2	28
3.4	Seleção: CP3+CP5 maior que 0,2	29
3.5	Matriz de autocorrelação entre os dez parâmetros no caso linear após a seleção CP3+CP5 maior que 0,2	29
3.6	Gráfico de dispersão entre componentes principais após seleção CP3+CP5 maior que 0,2	30
3.7	Seleção: 33CP3-CP1 maior que -3,3 e CP3 maior que -0,15	31
3.8	Gráfico de dispersão entre componentes principais após seleção 33CP3-CP1 maior que -3,3 e CP3 maior que -0,15	31
3.9	Matriz de correlação entre os doze parâmetros no caso do modelo quadrático	32
3.10	Gráfico de dispersão entre componentes principais para o caso do modelo quadrático	34
4.1	Histograma do <i>redshift</i> espectroscópico de estrelas e quasares do SDSS . .	36
4.2	Gráficos para QSO1	37
4.3	Gráficos para QSO2	38

4.4	Gráficos para QSO3	38
4.5	Histogramas do erro absoluto para média e mediana	39
4.6	Distribuição do desvio padrão do <i>redshifts</i> espectroscópico para os 10 e 100 objetos com menores RMSE	40
4.7	Distribuição do erro absoluto antes e após o refinamento	43
4.8	Distribuição do desvio padrão para o <i>redshift</i> dos 10 objetos com menores RMSE antes e após o refinamento.	43

Lista de Tabelas

3.1	Parâmetros das Componentes Principais (modelo linear)	28
3.2	Parâmetros das Componentes Principais (modelo quadrático)	33
4.1	Tabela de proporções dos erros absolutos do <i>redshift</i> de 80 quasares	40
4.2	Tabela de proporções dos quasares que apresentaram diferença na estimativa do <i>redshift</i> após o corte em z	41
4.3	Tabela de proporções dos erros absolutos do <i>redshift</i> de 80 quasares após corte em z	42

Sumário

1. Introdução	19
2. Base de dados	23
3. Desenvolvimento I	25
3.1 Análise de Componentes Principais (ACP) utilizando modelo linear	25
3.2 Análise de Componentes Principais (ACP) utilizando modelo quadrático .	31
4. Desenvolvimento II	35
4.1 Método de estimação do <i>redshift</i>	35
4.2 Exemplos de aplicação	36
4.3 Avaliação do método	38
4.4 Refinamento	42
5. Conclusões	45
Referências	47

Introdução

Galáxias com núcleos ativos (AGN) são caracterizadas por ter forte emissão de energia, supostamente ocasionada pela interação de um buraco negro central com a matéria. Uma classe de AGN são os quasares (QSO), objetos compactos que são as fontes mais luminosas do Universo, chegando a emitir até um trilhão de vezes mais energia que o Sol (Oliveira, 2014). Estes objetos são de grande interesse para a Cosmologia, dado que os maiores *redshifts* já observados correspondem aos quasares. O quasar com maior *redshift* já descoberto até o momento da publicação do artigo de Momjian et al. (2013) é o ULAS J112001.48+064124.3, com $z = 7,085$.

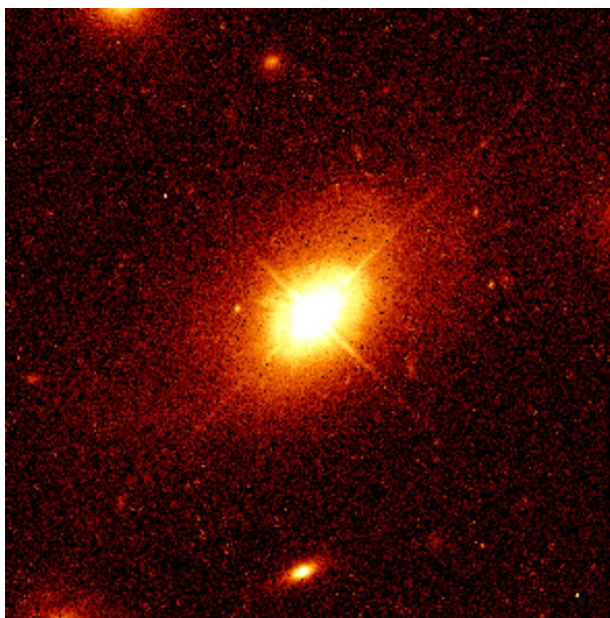


Figura 1.1: Imagem do QSO PHL 909 com $z = 0,171$, feita pelo Hubble Space Telescope na banda V. Créditos: John Bahcall (Institute for Advanced Study, Princeton) Mike Disney (University of Wales) e NASA/ESA.

Quasares emitem em todos os comprimentos de onda e possuem um espectro no visível tipicamente azul com linhas de emissão (Schneider, 2006). Quando observados no óptico possuem uma aparência pontual, o que dificulta a classificação entre estrela ou quasar por processamento de imagem.

O objetivo deste trabalho é utilizar métodos estatísticos e computacionais em observações do ALHAMBRA *Survey* [<http://www.alhambrasurvey.com/>], de forma que nos permita distinguir estas duas classes de objetos, possibilitando a descoberta de novos quasares.

Entre 2005 a 2012, o ALHAMBRA *Survey* observou com um telescópio de 3,5m no Calar Alto Observatory (CAHA, Espanha) uma área total de $4,0 \text{ deg}^2$ em oito regiões separadas do céu. Este *survey* cobre toda a região do visível ($3500\text{\AA}$ a $9700\text{\AA}$) com 20 bandas contínuas de largura aproximadamente $300\text{\AA}$ e a região do infravermelho próximo com as bandas JHKs, vide Figura 1.2 de Molino et al. (2013).

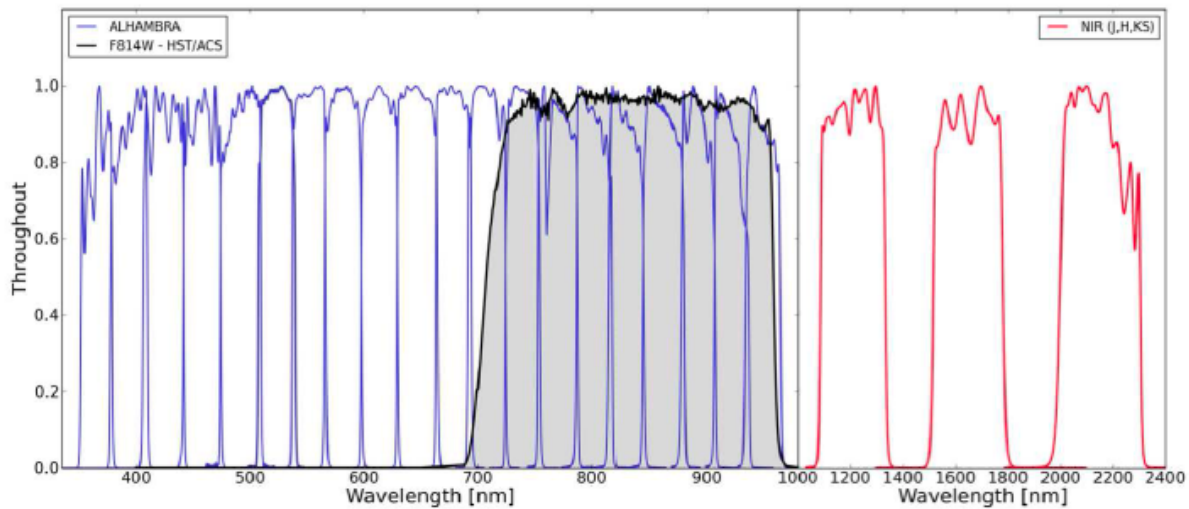


Figura 1.2: Conjunto de filtros do ALHAMBRA *survey*. À esquerda, as linhas contínuas em azul correspondem ao sistema óptico de filtros com aproximadamente $300\text{\AA}$ de largura. A linha em preto corresponde ao filtro sintético F814W. À direita, as linhas contínuas em vermelho representam as bandas em infravermelho próximo. (Molino et al., 2013)

Após obter uma rotina de separação de estrelas e quasares, nosso intuito é aplicá-lo em dados de novos *surveys* como o J-PAS [<http://www.j-pas.org/>]. A partir de 2015, em um período aproximado de 4 anos o J-PAS irá observar uma área de mais de 8000 deg^2 com um telescópio de 2,5m no Javalambre Observatory na Espanha (Benítez et al., 2014). Ao contrário do ALHAMBRA *Survey*, o J-PAS não possuirá filtros em infravermelho, tendo

apenas as 56 bandas no visível (Figura 1.3).

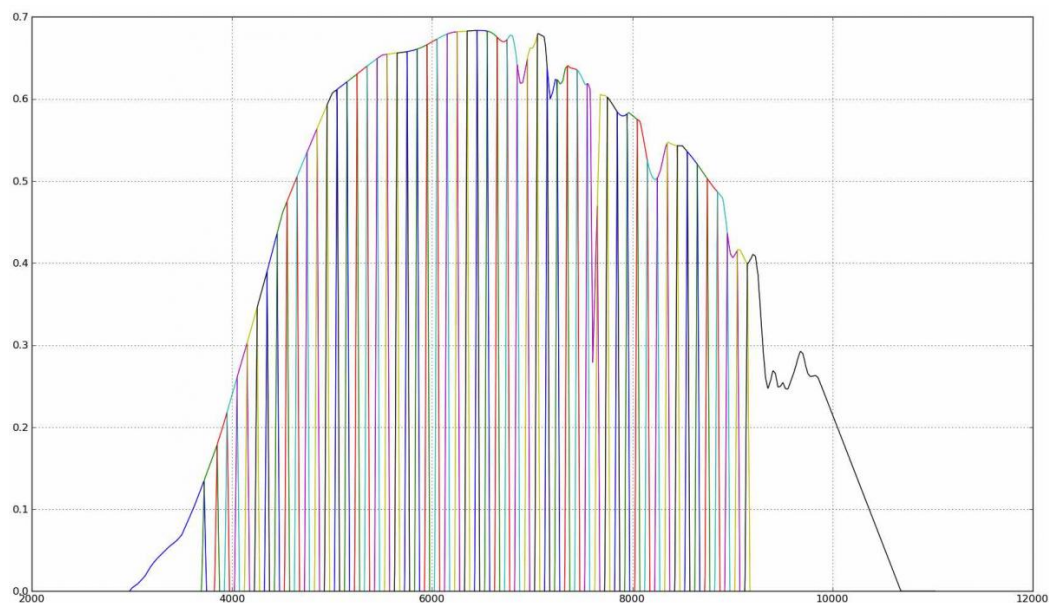


Figura 1.3: Conjunto de filtros J-PAS com 56 bandas no visível de aproximadamente 100Å de largura. Fonte: <http://jpas.astro.ufsc.br/pt-br/content/17>

No Capítulo 2 descrevemos os dados obtidos do ALHAMBRA *Survey* e o método de amostragem de estrelas e quasares. No Capítulo 3 discutimos a Análise de Componentes Principais que foram aplicados aos dados e seus resultados. Após obter os candidatos à quasar, elaboramos e testamos um método de estimação do *redshift* desses objetos, descrito no Capítulo 4.

Base de dados

A amostragem de estrelas e quasares foi feita a partir dos dados do *ALHAMBRA Survey*, que contém, entre outras, as seguintes informações:

- RA: Ascensão reta em graus
- DEC: Declinação em graus
- F365W, F396W, F427W, F458W, F489W, F520W, F551W, F582W, F613W, F644W, F675W, F706W, F737W, F768W, F799W, F830W, F861W, F892W, F923W, F954W: Magnitude total para cada um dos filtros, em que a numeração corresponde ao comprimento de onda (Por exemplo, F365W corresponde a 3650Å)
- J, H, KS: Magnitude total em infravermelho próximo
- F814W: Magnitude total para este filtro, em que as imagens analisadas foram geradas sinteticamente, a partir de uma função que relaciona alguns dos filtros anteriores
- Stellar_Flag: Fator que discrimina o objeto estatisticamente em galáxia (0), estrela (1) e desconhecido (0,5)

Outras variáveis contidas nos catálogos do *ALHAMBRA Survey* estão descritas em <http://cosmo.iaa.es/content/ALHAMBRA-Gold-catalog%20>.

O grupo de estrelas foi determinado a partir dos parâmetros $Stellar_Flag > 0,7$ e $17 < F814W < 22$. Note que, apesar do nome, este grupo pode conter quasares, dado que baseamos a amostragem pela sua estelaridade (Stellar.Flag).

Para amostrar o grupo de quasares, utilizamos o NED [<https://ned.ipac.caltech.edu/>], a partir da localização em ascensão reta (RA) e declinação (DEC) de cada apontamento

do ALHAMBRA *Survey* em um raio de 10,6066 arcmin. Foram encontrados um total de 160 quasares.

Após adquirir os dados, foi necessário fazer o cruzamento entre os objetos encontrados no NED com os objetos presentes nos catálogos do ALHAMBRA *Survey*. Para tanto, utilizamos um script em Python, encontrando um total de 82 correspondências.

Na Figura 2.1 é possível observar dois dos objetos identificados como quasares em uma imagem.fit proveniente do ALHAMBRA *Survey*, utilizando o *software* SAOImage DS9. Esta verificação visual foi conveniente para conferir se o tipo de objeto classificado como quasar condiz com o esperado. Como os objetos encontrados são pontuais, não há evidências para possíveis erros.

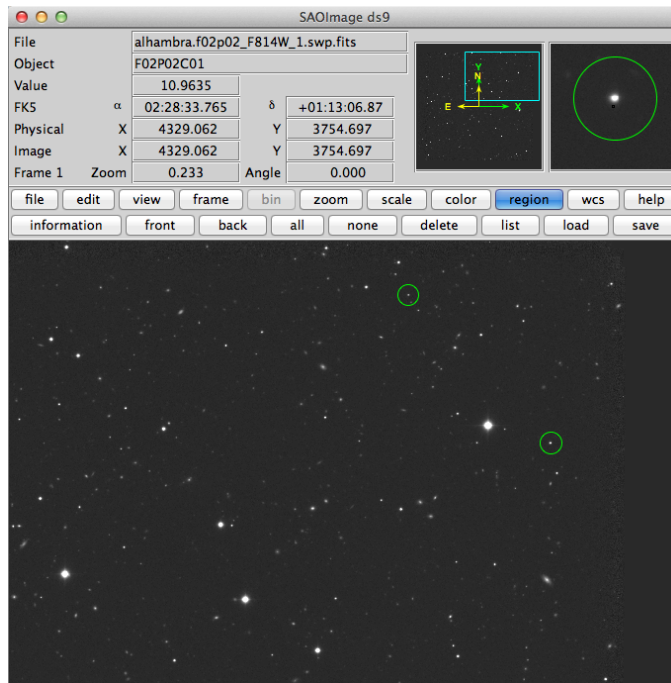


Figura 2.1: Localização dos quasares encontrados pelo NED em uma das imagens .fit do ALHAMBRA Survey, utilizando o *software* SAOImage DS9.

No Capítulo 3 descrevemos a Análise de Componentes Principais que foi aplicada aos dados. Para esta análise, selecionamos todas as estrelas do ALHAMBRA Survey seguindo os critérios citados anteriormente, totalizando 14745 estrelas, das quais consideramos as 10734 estrelas que não possuem dados faltantes. Os quasares que também foram selecionados por estes critérios foram retirados do grupo de estrelas e apenas 2 objetos foram retirados por falta de informação, totalizando 80 quasares. Portanto, a análise foi feita sobre uma amostra de 10814 objetos.

Desenvolvimento I

Em busca de um método para distinguir o grupo de estrelas do grupo de quasares, aplicamos a Análise de Componentes Principais (ACP). Nesta análise, são calculadas componentes principais a partir da combinação linear de n variáveis. As componentes principais representam um novo sistema de coordenadas não-correlacionadas na direção da máxima variabilidade dos dados (Johnson e Wichern, 2007). Assim, a ACP possibilita não apenas a redução de dimensão, mas também facilita a interpretação dos dados.

A ideia de usar a ACP no contexto de separação de estrelas e quasares é aplicá-la sobre alguns parâmetros que, *a priori*, sabemos ter comportamentos diferentes nos dois grupos. Após obter as componentes principais, esperamos que gráficos entre elas mostre regiões diferentes de concentração dos dois tipos de objetos. Na Seção 3.1 é descrita a ACP utilizando modelo linear e na Seção 3.2 descrevemos um modelo quadrático.

3.1 Análise de Componentes Principais (ACP) utilizando modelo linear

Os dez parâmetros que foram utilizadas na ACP são:

1. Desvios padrão dos resíduos
2. Assimetrias dos resíduos
3. Curtose dos resíduos
4. Coeficientes angulares dos modelos lineares
5. $H_\alpha = F644W - 2,5 \times \log_{10}(10^{\frac{-F613W}{2,5}} + 10^{\frac{-F675W}{2,5}})$

6. $H_\beta = F489W - 2,5 \times \log_{10}(10^{\frac{-F458W}{2,5}} + 10^{\frac{-F520W}{2,5}})$
7. $D_{4000} = F427W - F365W$
8. $J - H$
9. $H - KS$
10. $J - F954W$

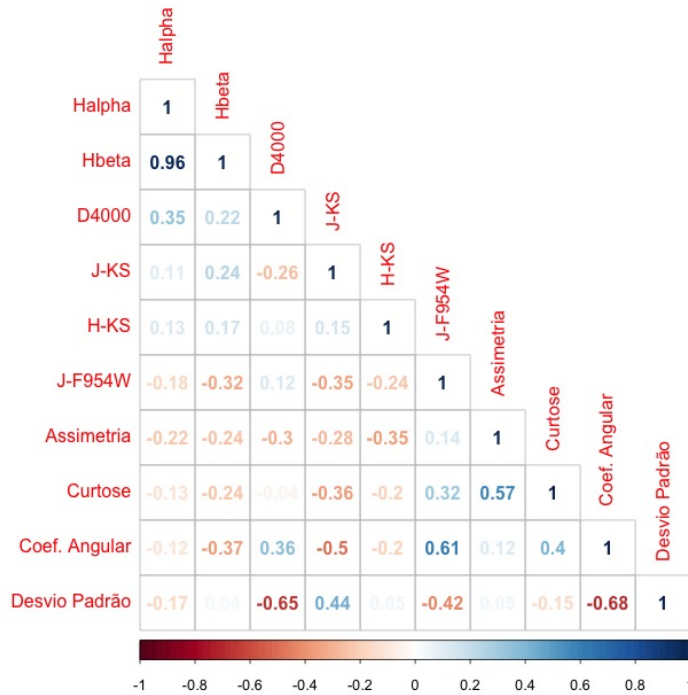


Figura 3.1: Matriz de autocorrelação entre os dez parâmetros.

Para ajustar um modelo linear de cada objeto, utilizamos a função $lm()$ do R, usando como variável dependente os valores dos filtros na faixa do visível e como variável explicativa, os valores de comprimento de onda. A partir do ajuste, obtivemos um vetor de resíduos, pelo qual foram encontrados os parâmetros de desvio padrão, assimetria e curtose. O parâmetro de coeficiente angular foi encontrado a partir dos objetos de saída da função

$lm()$ e os parâmetros restantes estão diretamente relacionados com os dados contidos nos catálogos. A matriz de autocorrelação entre os parâmetros está na Figura 3.1.

Utilizando a função *prcomp()* do R, geramos as componentes principais e gráficos de dispersão entre elas foram plotados no TOPCAT, a fim de verificar o comportamento das estrelas e dos quasares. Desta forma, seria possível avaliar qual componente melhor discrimina os grupos e selecionar novos candidatos a quasar.

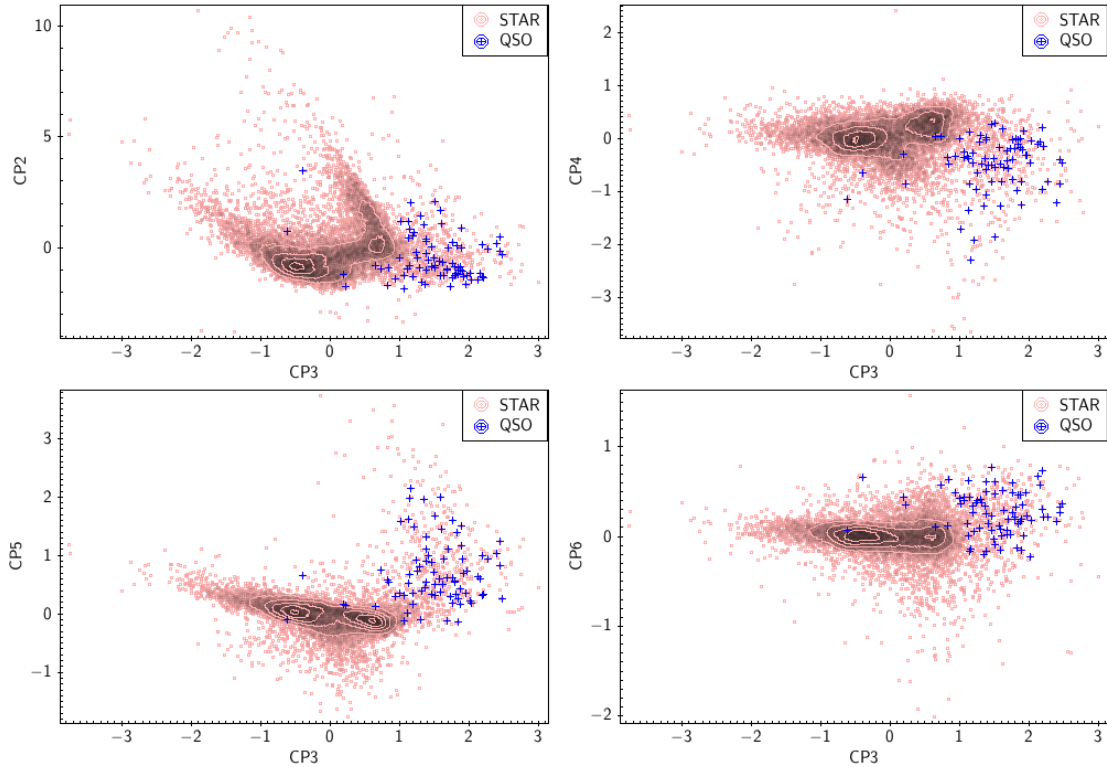


Figura 3.2: Gráfico de dispersão entre componentes principais, com símbolos azuis e rosas representando quasares e estrelas, respectivamente. (a) À esquerda no canto superior, gráfico entre CP3 e CP2. (b) À direita no canto superior, gráfico entre CP3 e CP4. (c) À esquerda no canto inferior, gráfico entre CP3 e CP5. (d) À direita no canto inferior, gráfico entre CP3 e CP6.

A componente principal que melhor discriminou os dois grupos foi a CP3 e alguns dos gráficos estão dispostos na Figura 3.2. Em todos os gráficos é possível notar uma certa tendência tanto para os quasares, quanto para as estrelas. Demais gráficos entre as demais componentes foram omitidos desta monografia devido ao grande número e por não haver uma distinção tão clara de regiões de estrelas e quasares como nos gráficos acima.

Notamos também que houve a distinção de três sequências de estrelas na Figura 3.2 (a), que pode ser melhor observado na Figura 3.3.

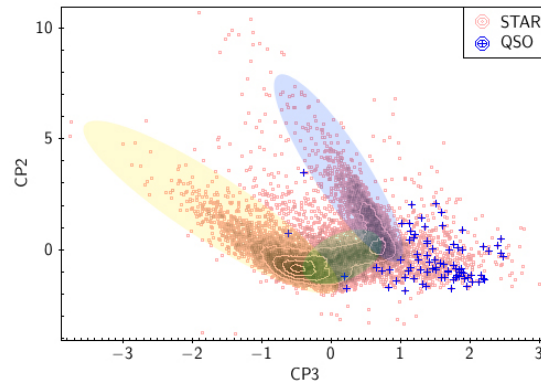


Figura 3.3: Gráfico de dispersão entre CP3 e CP2. É possível notar três sequências de estrelas salientadas em amarelo, verde e azul

Obtivemos os resultados contidos na Tabela 3.1, em que é possível notar que a primeira componente principal já representa mais de 90% da variabilidade total dos dados.

Tabela 3.1 - Parâmetros das Componentes Principais (CP) com relação à amostra reduzida de estrelas.

	Desvio Padrão	Proporção da Variância	Proporção acumulada
CP1	4,774	0,9018	0,9018
CP2	1,266	0,0634	0,9652
CP3	0,719	0,0204	0,9857
CP4	0,391	0,0060	0,9917
CP5	0,350	0,0048	0,9966
CP6	0,195	0,0015	0,9981
CP7	0,162	0,0010	0,9991
CP8	0,130	0,0007	0,9998
CP9	0,074	≈ 0	≈ 1
CP10	0,001	≈ 0	1

Com o intuito de melhorar o resultado obtido, selecionamos os objetos para os quais $CP3 + CP5 > 0,2$, cuja determinação foi feita visualmente, vide Figura 3.4. Nesta seleção mantivemos incluídos todos os quasares.

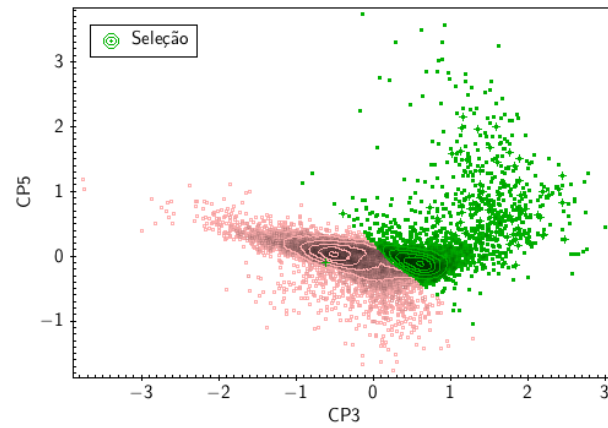


Figura 3.4: Os objetos selecionados para uma nova ACP estão representados com a cor verde no gráfico CP3xCP5.

Uma segunda ACP foi realizada e obtivemos que a CP3 desta nova análise é a componente que melhor discriminou os dois grupos, como pode ser observado na Figura 3.6.

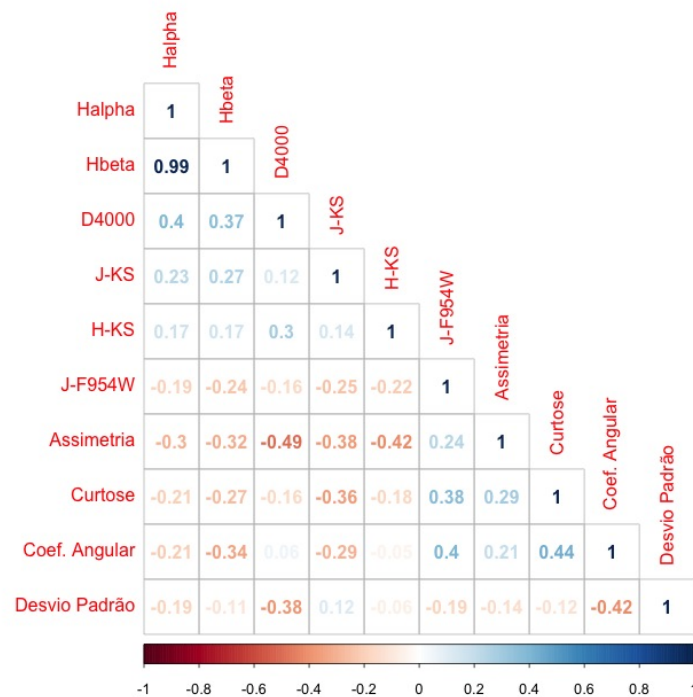


Figura 3.5: Matriz de autocorrelação entre os dez parâmetros dos objetos selecionados.

Observamos na Figura 3.6 (a) uma região de quasares e uma sequência muito clara de estrelas, que provavelmente se trata de anãs brancas do tipo D_a , devido à alta correlação entre H_α e H_β , vide Figura 3.5.

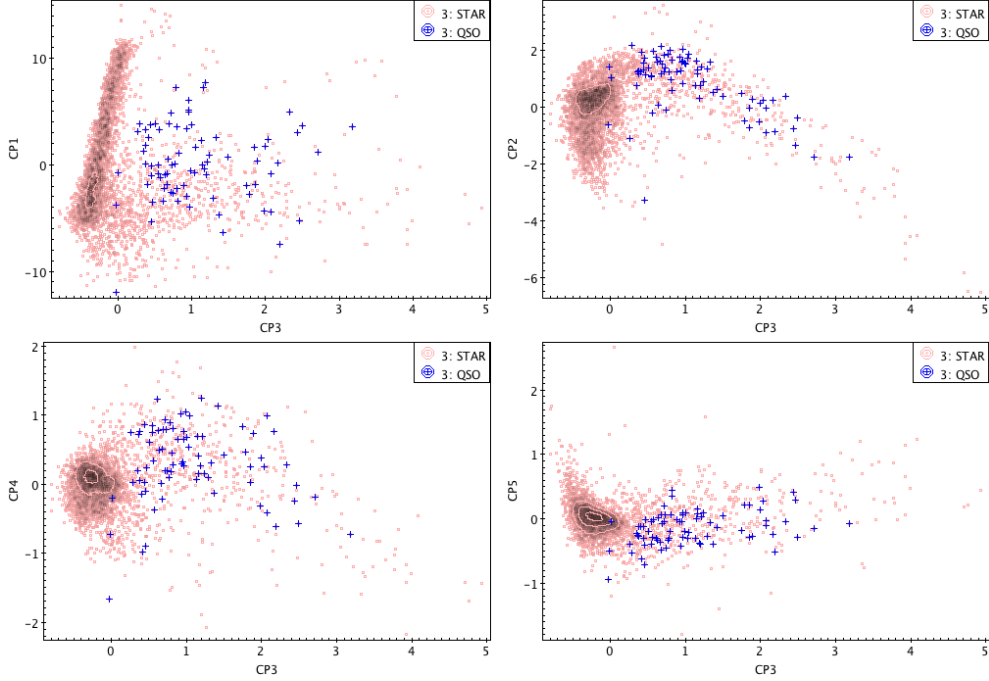


Figura 3.6: Gráfico de dispersão entre componentes principais, com símbolos azuis e rosas representando quasares e estrelas, respectivamente. (a) À esquerda no canto superior, gráfico entre CP3 e CP1. (b) À direita no canto superior, gráfico entre CP3 e CP2. (c) À esquerda no canto inferior, gráfico entre CP3 e CP4. (d) À direita no canto inferior, gráfico entre CP3 e CP5.

Seria então viável tentar fazer uma nova seleção de objetos desconsiderando essa sequência de estrelas, com os seguintes critérios determinados visualmente: $33CP3 - CP1 > -3,3$ e $CP3 > -0,15$ (Figura 3.7). Os dois gráficos da Figura 3.8 mostram que foi possível melhorar a representação gráfica ao refazer a análise sem os objetos que podem ser consideradas como estrelas. Além disso, como os quasares e as estrelas encontram-se numa mesma região do gráfico, podemos admitir que essas estrelas são, na verdade, candidatas a quasar, com um total de 985 objetos.

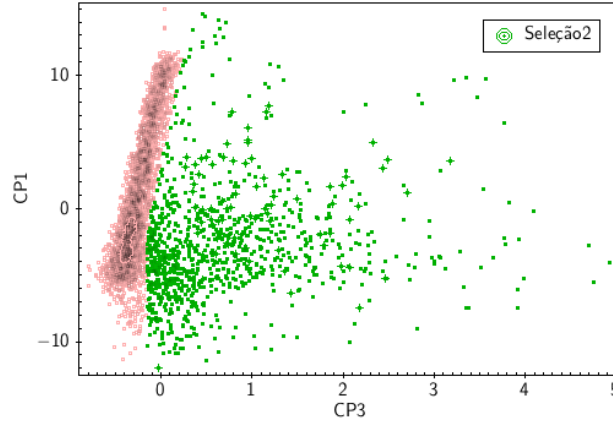


Figura 3.7: Os objetos selecionados para uma terceira ACP estão representados com a cor verde no gráfico CP3xCP1.

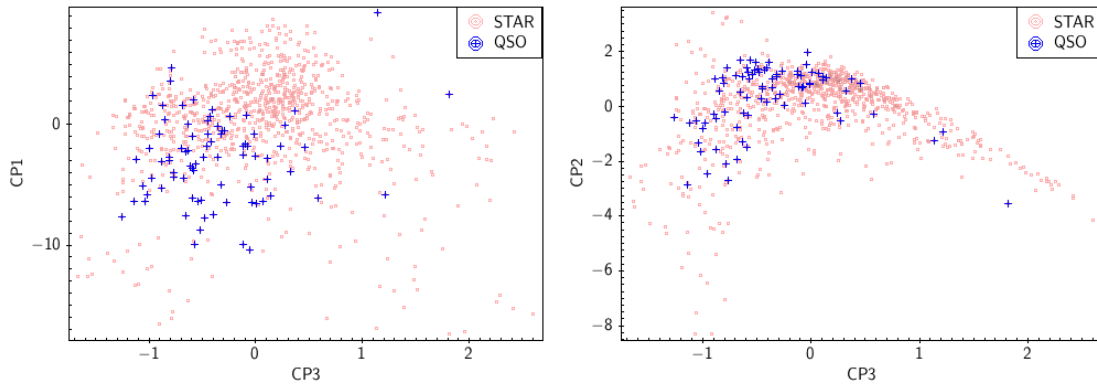


Figura 3.8: Gráfico de dispersão entre componentes principais, com símbolos azuis e rosas representando quasares e estrelas, respectivamente. (a) À esquerda, gráfico entre CP3 e CP1. (b) À direita, gráfico entre CP3 e CP2.

3.2 Análise de Componentes Principais (ACP) utilizando modelo quadrático

Uma segunda abordagem de ACP foi realizada, trocando os modelos lineares por modelos quadráticos e utilizando a mesma função $lm()$ do R, em que a mesmas variáveis dependente e explicativa foram usadas, porém com um termo quadrático adicionado.

Consideramos os mesmos parâmetros que foram utilizados na ACP da seção anterior, porém com os parâmetros 1 a 4 relacionados ao modelo quadrático. Além destes, outros dois parâmetros foram adicionados:

11. Curvatura do espectro ou coeficiente do termo quadrático

12. $B_{jump} = F396W - F365W$

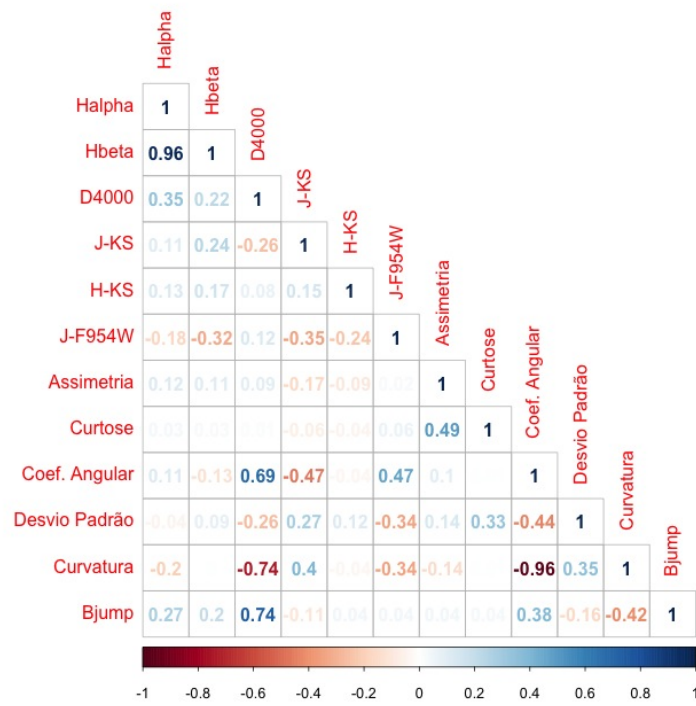


Figura 3.9: Matriz de autocorrelação entre os doze parâmetros.

Desta forma, temos no total 12 parâmetros, cujas correlações estão dispostas na Figura 3.9. Pela Tabela 3.2 notamos que as duas primeiras componentes representam mais de 90% da variabilidade total dos dados.

Tabela 3.2 - Parâmetros das Componentes Principais (CP) com relação à amostra reduzida de estrelas.

	Desvio Padrão	Proporção da Variância	Proporção acumulada
CP1	4,772	0,8883	0,8883
CP2	1,281	0,0640	0,9523
CP3	0,782	0,0238	0,9762
CP4	0,496	0,0096	0,9858
CP5	0,465	0,0085	0,9942
CP6	0,241	0,0023	0,9965
CP7	0,201	0,0016	0,9981
CP8	0,162	0,0010	0,9991
CP9	0,134	0,0007	0,9998
CP10	0,066	0,0001	≈ 1
CP11	0,003	≈ 0	≈ 1
CP12	≈ 0	≈ 0	1

Novamente fizemos a análise dos gráficos de dispersão entre componentes principais e notamos que a componente que melhor discriminou as regiões de quasares e estrelas foi a terceira componente. Alguns destes gráficos estão dispostos na Figura 3.10.

Esta abordagem não foi aprofundada, como no caso linear descrito na seção anterior. Contudo, ainda é possível notar que as componentes desta análise distinguem bem as duas regiões e que pode ser uma alternativa à análise feita anteriormente.

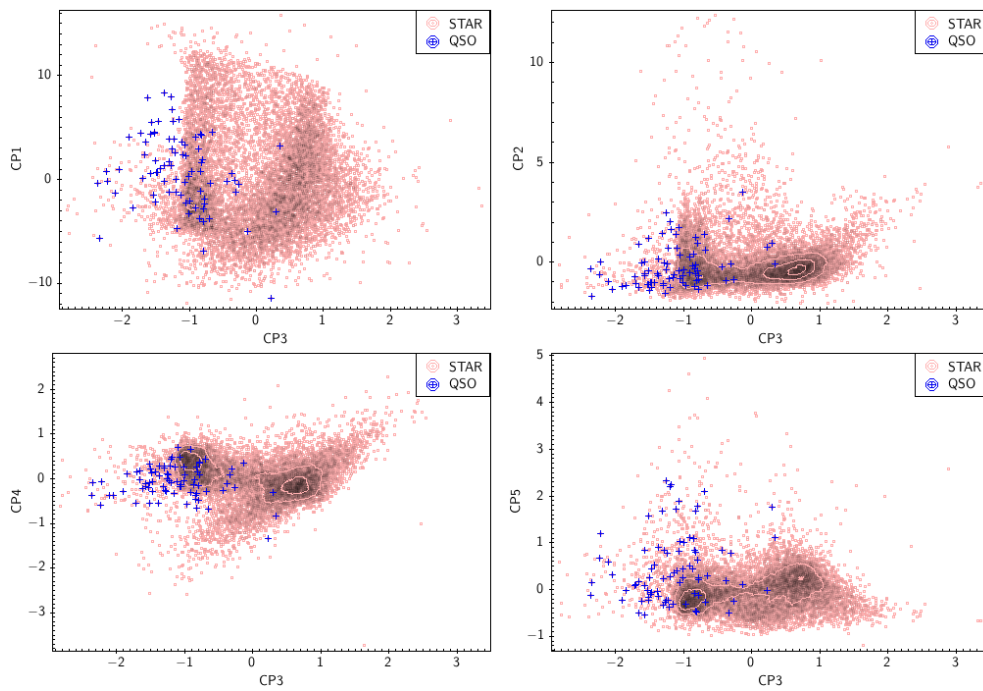


Figura 3.10: Gráfico de dispersão entre componentes principais para o caso quadrático, com símbolos azuis e rosa representando quasares e estrelas, respectivamente. (a) À esquerda no canto superior, gráfico entre CP3 e CP1. (b) À direita no canto superior, gráfico entre CP3 e CP2. (c) À esquerda no canto inferior, gráfico entre CP3 e CP4. (d) À direita no canto inferior, gráfico entre CP3 e CP5.

Desenvolvimento II

Após a aplicação da ACP nos dados do ALHAMBRA Survey e a consequente obtenção de 985 candidatos à quasar descritos na Seção 3.1, foi necessário desenvolver um método de estimação dos seus *redshifts* para que seja possível confirmar a classificação destes objetos como quasares, já que o *redshift* de um quasar é significativamente maior que de uma estrela (Oliveira, 2014). Na Seção 4.1 descrevemos este método e nas Seções 4.2 e 4.3 apresentamos alguns resultados de sua aplicação em quasares conhecidos do ALHAMBRA Survey. Por fim, na Seção 4.4 descrevemos e avaliamos um processo de refinamento do método.

4.1 Método de estimação do *redshift*

O método elaborado para estimar o *redshift* dos candidatos à quasar segue os passos:

1. Obter uma amostra de espectros de quasares e estrelas do SDSS (*Sloan Digital Sky Survey*) [<http://www.sdss.org/>] convoluídos pelos filtros do ALHAMBRA Survey
2. Ajustar uma reta para cada quasar do SDSS e obter os resíduos do ajuste. Realizar o mesmo procedimento para os objetos do ALHAMBRA Survey
3. Para cada candidato do ALHAMBRA Survey, calcular o RMSE (Root Mean Square Error) entre o conjunto de resíduos desse candidato e o conjunto de resíduos de todos os objetos do SDSS. Supondo que $x = (x_1, \dots, x_n)$ denota o conjunto de resíduos de um objeto no ALHAMBRA Survey e $x' = (x'_1, \dots, x'_n)$, o conjunto de resíduos de um objeto do SDSS, o RMSE é calculado por:

$$RMSE = \sqrt{\frac{\sum (x_i - x'_i)^2}{n}} \quad (4.1)$$

4. Ordenar os objetos do SDSS segundo o RMSE para cada objeto do ALHAMBRA *Survey*
5. Determinar o *redshift* do objeto do ALHAMBRA *Survey* a partir de uma estatística dos *redshifts* espectroscópicos do SDSS dos objetos com menores valores de RMSE

Obtivemos uma amostra de 90702 espectros do SDSS convoluídos pelos filtros do ALHAMBRA *Survey*, sendo 78962 quasares e 11740 estrelas. A distribuição do *redshift* espectroscópico destes objetos pode ser visualizada na Figura 4.1.

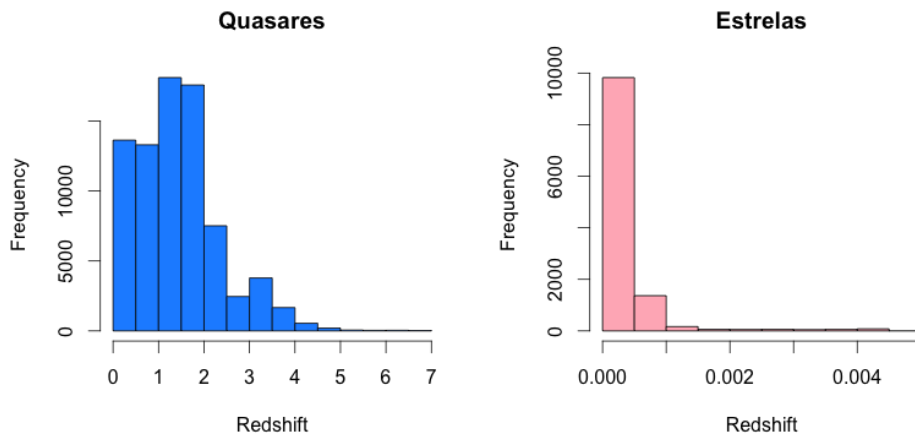


Figura 4.1: Histograma para 78962 e 11740 *redshifts* espectroscópicos da amostra de quasares (à esquerda) e estrelas (à direita) do SDSS, respectivamente.

Assim como a aplicação da ACP, os passos de 2 a 5 foram codificadas na linguagem R. Para o ajuste linear dos objetos do SDSS e do ALHAMBRA *Survey* foram considerados apenas os filtros dentro do escopo da faixa de comprimento de onda de observação do SDSS. Logo, foram desconsideradas as observações nos filtros F365W, F814W e F954W.

4.2 Exemplos de aplicação

Testamos o método descrito na Seção 4.1 em 25 quasares conhecidos do ALHAMBRA *Survey*, a fim de avaliar os diversos casos que podem ocorrer. Para tal, foram feitos gráficos de dispersão do *redshift* espectroscópico pelo RMSE e um histograma para o *redshift* espectroscópico, sendo ambos os gráficos referentes aos 100 objetos com menores RMSE. Apresentamos nesta seção três principais casos.

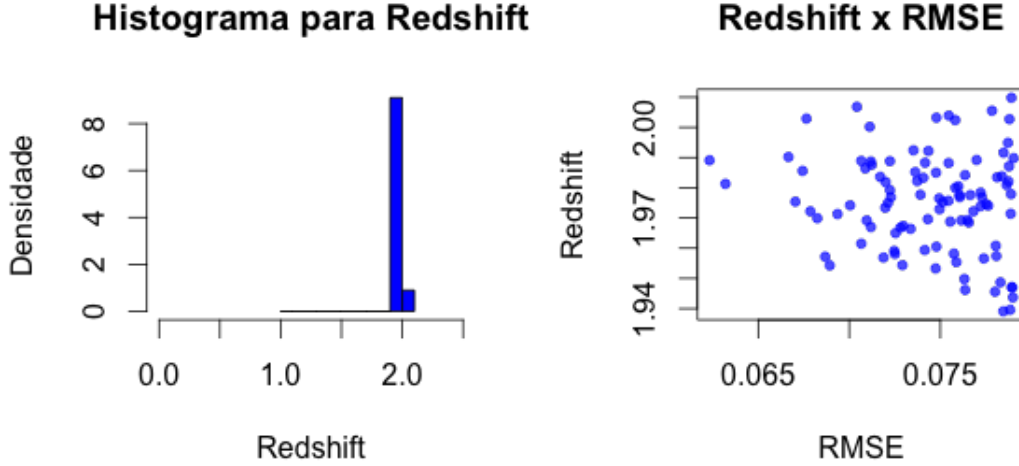


Figura 4.2: Resultados gráficos para o QSO₁ referentes aos 100 objetos do SDSS com menores RMSE. (a) À esquerda, histograma do *redshift* espectroscópico. (b) À direita, gráfico de dispersão entre RMSE e *redshift* espectroscópico.

Os resultados presentes na Figura 4.2 para o QSO₁ representam uma situação ideal, a qual esperamos que o método retorne para os demais quasares conhecidos do ALHAMBRA *Survey* e também para os quasares dentre os 985 candidatos. No exemplo da Figura 4.2, os 100 menores RMSE possuem uma média para o *redshift* $\bar{z}_{100} = 1,976$ com um desvio padrão $dp(\bar{z}_{100}) = 0,016$. O valor conhecido de *redshift* para este quasar obtido pelo NED é $z_{NED} = 1,962374$.

Já para o QSO₂, cujo valor conhecido de *redshift* é $z_{NED} = 1,524857$, é possível perceber dois picos de *redshift* pela Figura 4.3(a). Além disso, há dois valores de *redshift* discrepantes em relação aos demais, iguais a 6,7870 e 4,8895, este que representa o objeto de menor RMSE. Nota-se também que o segundo maior pico da Figura 4.3(a) refere-se ao intervalo $0 < z < 0,1$, no qual a mediana do *redshift* é de 0,0002. Este pico de *redshift* seria incompatível para um quasar.

Para o QSO₃ com $z_{NED} = 0,458619$, cujos gráficos de histograma e de dispersão estão contidos na Figura 4.4, o resultado do método não foi satisfatório, dado que é possível identificar ao menos três picos de possíveis valores de z . Além disso, o maior pico de *redshift* está em torno de 1 e apenas o terceiro maior pico está próximo ao valor de z_{NED} .

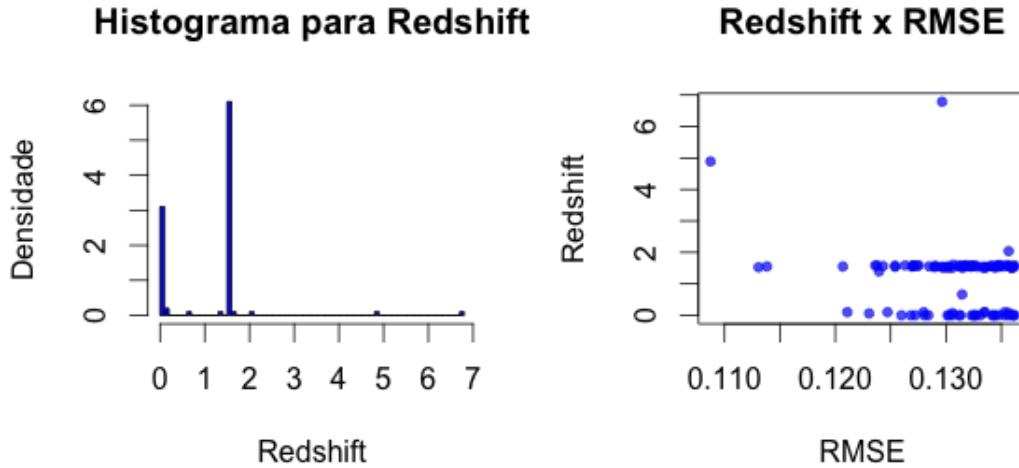


Figura 4.3: Resultados gráficos para o QSO₂ referentes aos 100 objetos do SDSS com menores RMSE. (a) À esquerda, histograma do *redshift* espectroscópico. (b) À direita, gráfico de dispersão entre RMSE e *redshift* espectroscópico.

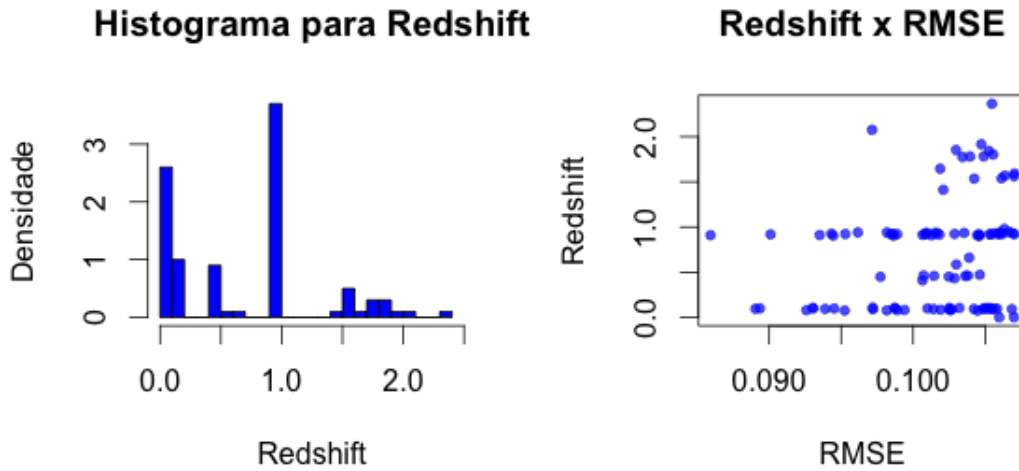


Figura 4.4: Resultados gráficos para o QSO₃ referentes aos 100 objetos do SDSS com menores RMSE. (a) À esquerda, histograma do *redshift* espectroscópico. (b) À direita, gráfico de dispersão entre RMSE e *redshift* espectroscópico.

4.3 Avaliação do método

A informação obtida do NED do *redshift* espectroscópico do grupo de 80 quasares conhecidos nos permite avaliar a acurácia do método. Para tanto, consideramos os estimadores: média dos 10 ($\bar{z}_{10}$) e dos 100 ($\bar{z}_{100}$) objetos com menores RMSE; mediana dos 10 ($m(z)_{10}$)

e dos 100 ($m(z)_{100}$) objetos com menores RMSE. Já o desvio padrão será conveniente para avaliar a precisão.

Após obter as estatísticas, calculamos o erro absoluto que é a diferença entre o z_{NED} e a estimativa. As distribuições do erro absoluto para $\bar{z}_{10}$, $\bar{z}_{100}$, $m(z)_{10}$ e $m(z)_{100}$ estão presentes na Figura 4.5, em que é possível perceber uma maior concentração de erros em torno do zero para todos os casos. Pela Tabela 4.1 vemos que pelo menos 70% quasares tiveram seus *redshifts* determinados com no máximo 0,01 de erro para todos os estimadores.

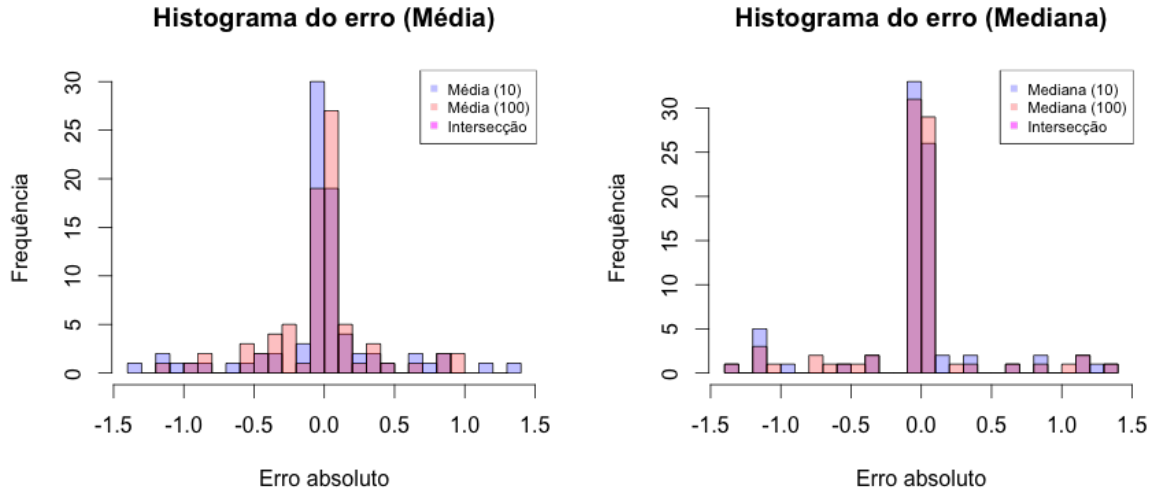


Figura 4.5: Distribuição do erro absoluto, ou seja, da diferença entre o *redshift* espectroscópico e a estimativa. (a) À esquerda temos as distribuições dos erros para a média dos 10 e dos 100 objetos com menores RMSE. (b) À direita temos as distribuições dos erros para a mediana dos 10 e dos 100 objetos com menores RMSE.

Observamos que 53,75% dos erros de $\bar{z}_{10}$ são maiores que os erros correspondentes de $m(z)_{10}$. Para $\bar{z}_{100}$, os erros são maiores que os erros de $m(z)_{100}$ em 60% dos casos. Isto indica que o uso da mediana como estimador pode ser mais eficaz que o uso da média.

Quando comparados os erros de $\bar{z}_{10}$ e $\bar{z}_{100}$, 55% dos *redshifts* dos quasares possuem maior erro ao calcular a estimativa com 100 objetos ao invés de apenas 10 objetos. Esta porcentagem sobre para 61,25% quando considerada a mediana. Estes resultados indicam o aumento da acurácia ao estimar o *redshift* a partir dos 10 objetos com menores RMSE.

Tabela 4.1 - Proporções dos erros absolutos de 80 quasares considerando a média e a mediana como estimadores do *redshift*.

	10 menores RMSE		100 menores RMSE	
	Média	Mediana	Média	Mediana
Erro > 1	2,5%	5%	0	5%
1 > Erro > 0,1	17,5%	11,25%	16,25%	7,50%
0,1 > Erro > 0,01	10%	15%	25%	13,75%
0,01 > Erro > 0,001	11,25%	18,75%	26,25%	20%
Erro < 0,001	58,75%	56,25%	47,5%	56,25%

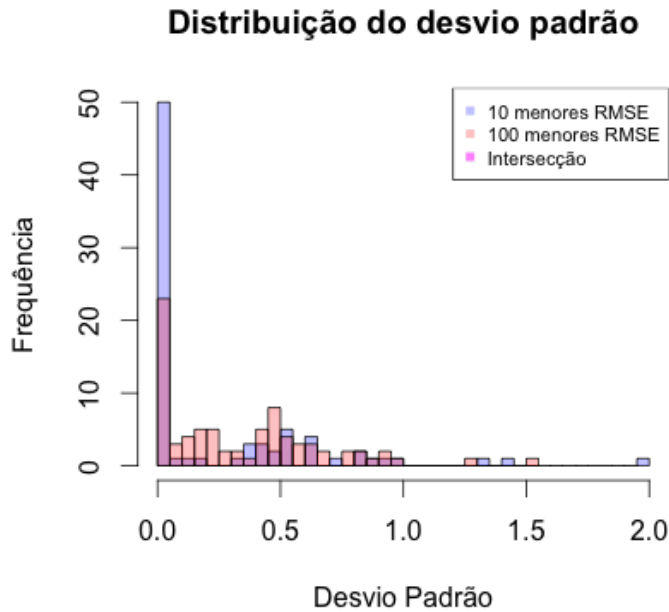


Figura 4.6: Distribuição do desvio padrão do *redshifts* espectroscópico para os 10 e 100 objetos com menores RMSE.

Um desvio padrão muito próximo de zero pode indicar casos em que o *redshift* está bem determinado como o QSO₁ (Figura 4.2), enquanto um desvio padrão alto pode indicar a existência de diversos picos de *redshift* como no caso do QSO₃ (Figura 4.4). A distribuição do desvio padrão do *redshift* espectroscópico contido na Figura 4.6 nos mostra que mais da metade dos quasares apresentaram boa precisão na determinação do *redshift* quando considerados apenas os 10 objetos com menores RMSE.

Na Seção 4.2 vimos o caso do QSO₂ em que há confusão na determinação do *redshift* com objetos classificados como estrelas no SDSS (Figura 4.3). Como os valores dos *redshifts* espectroscópicos das estrelas são muito baixas (Figura 4.1), decorreria uma subestimação do *redshift* pelo método. Verificamos que o valor máximo de *redshift* espectroscópico para as estrelas é de 0,00046 e reaplicamos o método considerando apenas os objetos com *redshifts* maiores que este valor. Um total de 13 objetos do SDSS tidos como quasares foi descartado por esta seleção.

As proporções de quasares do ALHAMBRA *Survey* que apresentaram diferença nas estatísticas depois do corte em $z > 0,00046$ feito no catálogo do SDSS encontram-se na Tabela 4.2. Em todos os casos que houve diferença nas estatísticas, houve o aumento da estimativa do *redshift* e diminuição do erro. Quando considerados apenas os 10 objetos com menores RMSE, apenas 12,50% das médias e 8,75% das medianas mudaram com o corte.

Tabela 4.2 - Proporção de quasares que apresentaram diferença após o corte em $z > 0,00046$ para cada estimador. As três últimas linhas da tabela indicam a proporção de quasares que apresentaram diferença a partir da unidade, da primeira e da segunda casa decimal em relação ao total de objetos que apresentaram alguma diferença.

	10 menores RMSE		100 menores RMSE	
	Média	Mediana	Média	Mediana
Houve diferença	12,50%	8,75%	37,50%	33,75%
Diferença a partir da unidade	10%	28,57%	0	11,11%
Diferença a partir da primeira casa decimal	100%	57,14%	26,67%	18,52%
Diferença a partir da segunda casa decimal	100%	96,67%	85,71%	33,33%

A Tabela 4.3 com resultados após o corte em $z > 0,00046$ nos mostra um aumento no percentual de quasares com erros de no máximo 0,001 em comparação à Tabela 4.1, indicando melhoria na estimação.

Tabela 4.3 - Proporções dos erros absolutos de 80 quasares considerando a média e a mediana como estimadores do *redshift* após o corte em $z > 0,00046$.

	10 menores RMSE		100 menores RMSE	
	Média	Mediana	Média	Mediana
Erro > 1	2,5%	3,75%	0	2,5%
1 > Erro > 0,1	12,5%	5%	13,75%	1,25%
0,1 > Erro > 0,01	11,25%	17,5%	23,75%	16,25%
0,01 > Erro > 0,001	11,25%	15%	8,75%	20%
Erro < 0,001	62,5%	58,75%	53,75%	60%

4.4 Refinamento

Uma forma de refinar o método consiste em criar novos espectros de forma a obter um intervalo mais abrangente de *redshift*. Ou seja, cada objeto do SDSS listado entre os 10 menores RMSE teve seu espectro deslocado dez vezes dentro de um intervalo do filtro do ALHAMBRA *Survey* que corresponde a 300 Å, resultando em novos espectros a cada 30 Å. Desta forma, o RMSE é calculado novamente para cada novo conjunto de resíduos.

Para tanto, foi necessário obter a identificação de todos os quasares do catálogo do SDSS e obter seu espectro original pela plataforma CasJobs [<http://skyserver.sdss.org/casjobs/>]. Cada espectro foi deslocado por um código elaborado em Python que gerou um novo catálogo com 965725 objetos. Após este processo, estes espectros foram convoluídos com os filtros do ALHAMBRA *Survey*, para possibilitar a obtenção dos resíduos do ajuste linear e o cálculo do RMSE.

Na Figura 4.7 temos as distribuições do erro absoluto para a estimativa do *redshift* feita pela média e pela mediana dos 10 objetos com menores RMSE antes e após o refinamento. Após o refinamento, houve uma diminuição do erro absoluto para 50% dos quasares quando considerado o estimador $\bar{z}_{10}$. Já para o estimador $m(z)_{10}$, houve melhora em apenas 31,25% dos dados.

A distribuição do desvio padrão do *redshift* dos 10 objetos com menores RMSE antes e após o refinamento encontra-se na Figura 4.8. Houve um aumento no desvio padrão em 60% dos quasares conhecidos do ALHAMBRA *Survey*.

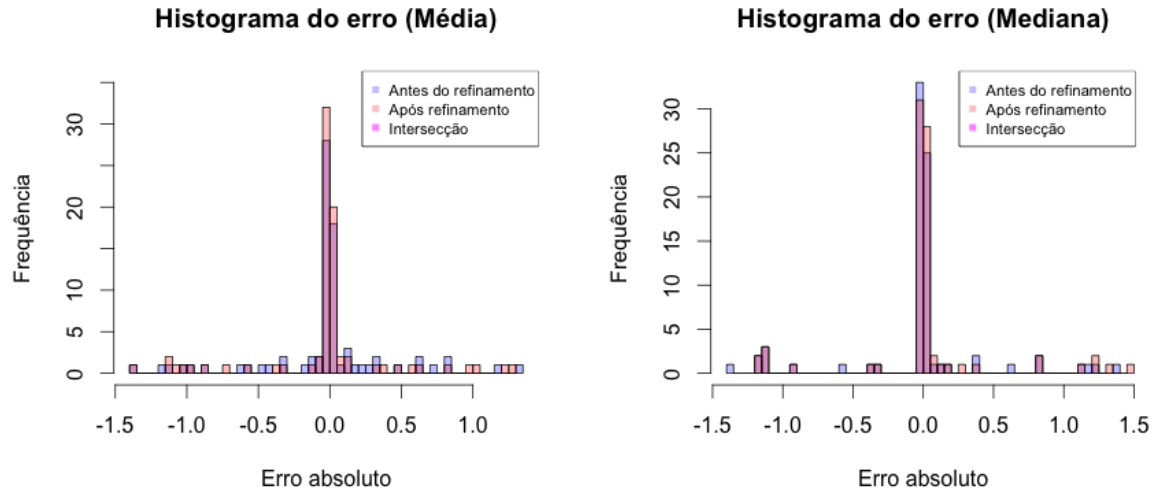


Figura 4.7: Distribuição da diferença (erro absoluto) entre o *redshift* espectroscópico e a estimava antes e após o refinamento. (a) À esquerda temos as distribuições dos erros para a média dos 10 objetos com menores RMSE. (b) À direita temos as distribuições dos erros para a mediana dos 10 objetos com menores RMSE.

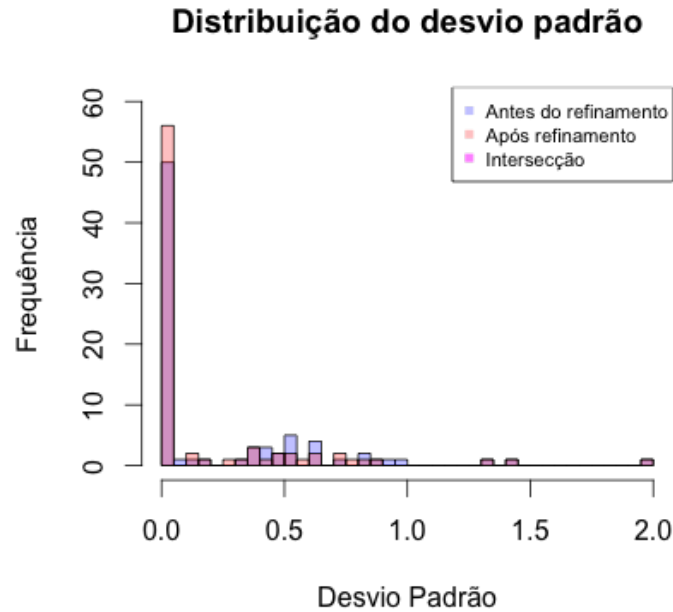


Figura 4.8: Distribuição do desvio padrão para o *redshift* dos 10 objetos com menores RMSE antes e após o refinamento.

Conclusões

Este projeto visa a separação de estrelas e quasares por meio de ferramentas estatísticas e computacionais, a fim de realizar a descoberta de novos quasares. Para tanto, determinamos dois conjuntos de parâmetros de distinção entre estrelas e quasares com base nos dados disponíveis do ALHAMBRA *Survey*, resultando em duas abordagens de análise de componentes principais. Nas duas abordagens foi possível distinguir regiões onde se encontram as estrelas e os quasares em gráficos de dispersão entre duas componentes principais. Por simplicidade, mantivemos o foco na análise do caso em que foi utilizado um modelo linear para determinação dos parâmetros de *input*.

Foi possível determinar 985 candidatos a quasar que *a priori* estavam classificados no grupo de estrelas. Os gráficos obtidos também mostram a possibilidade de estudar e classificar as sequências de estrelas que apareceram, podendo gerar resultados paralelos para a área de Astrofísica Estelar.

Para que seja possível verificar se os candidatos são realmente quasares, elaboramos um método para estimar seus *redshifts*. Calculamos o RMSE a partir dos resíduos de um ajuste de modelo linear das observações do ALHAMBRA *Survey* e do espectro convoluído do SDSS e estimamos o *redshift* pelos 10 e 100 objetos com menores valores de RMSE.

Resultados presentes na Seção 4.3 nos mostram que a estimação do *redshift* por estimadores como média e mediana pode ser eficiente, principalmente usando a mediana dos *redshifts* espectroscópicos dos 10 objetos com menores RMSE e considerando o corte $z > 0,00046$. Porém, o refinamento do método descrito na Seção 4.4 além de demandar um elevado tempo operacional, não se mostrou tão eficiente quanto se era esperado, o que nos levou à decisão de não aplicá-lo.

Deste projeto concluímos que a ACP pode ser viável para descobrir novos quasares e

que o método elaborado de estimação do *redshift* possibilitará a classificação dos candidatos à quasar, dado que obtivemos bons resultados mesmo ainda sendo necessário aperfeiçoá-lo.

Referências Bibliográficas

- Benítez N., Dupke R., Moles M., Sodre L., Cenarro J., Marin-Franch A., Taylor K., Cristóbal D., Fernandez-Soto A., Mendes de Oliveira C., Cepa-Nogue J., Abramo L. R., Alcaniz J. S., Overzier R., et al. J-PAS: The Javalambre-Physics of the Accelerated Universe Astrophysical Survey, JPAS "Red Book", 2014
- Johnson R. A., Wichern D. W., Applied Multivariate Statistical Analysis. Pearson, 2007
- Molino A., Benítez N., Moles M., Fernández-Soto A., Cristóbal-Hornillos D., Ascaso B., Jiménez-Teja Y., Schoenell W., Arnalte-Mur et al. The ALHAMBRA Survey: Bayesian Photometric Redshifts with 23 bands for 3 squared degrees, Monthly Notices of the Royal Astronomical Society, 2013
- Momjian E., Carilli C. L., Walter F., Venemans B., The Highest Redshift Quasar at $z=7.085$: A Radio Quiet Source, AJ, 2013
- Oliveira K. S. e Saraiva M. F., Astronomia e Astrofísica. Editora Livraria da Física, 2014, 760 p.
- Schneider P., Extragalactic Astronomy and Cosmology. Springer, 2006