

Spectrum Classification of the Quasars from SDSS with the Redshift $z \sim 3$ with PCA

Zeynab Kiamehr¹

¹ Basic Sciences group, Department of Marine Sciences, Chabahar Maritime University, Chabahar, Iran

Correspondence: Zeynab Kiamehr, Basic Sciences group, Department of Marine Sciences, Chabahar Maritime University, Chabahar, Iran. Tel: 98-054-35320020-3. E-mail: z.kiamehr@cmu.ac.ir

Received: June 26, 2015

Accepted: September 3, 2015

Online Published: October 31, 2015

doi:10.5539/mas.v9n12p99

URL: <http://dx.doi.org/10.5539/mas.v9n12p99>

Abstract

The spectrum lines of different quasars (QSOs) have been investigated, emphasis on the weak emission lines existent in the Ly α forest region using principal component analysis method (PCA) in the wavelength range from 1020 to 1600Å. The first, the second and the third principal component spectra (PCS) involve 63.4, 14.5 and 6.2% of the variance respectively, as the first seven PCS involve 96.1% of the total variance. The first PCS contain peak from high ionization emission lines namely the emission lines of Ly α and Ly β (OVI, NV, SiIV, and CIV), these peaks are sharp and strong and the second PCS has peaks from low ionization emission lines (FeII, FeIII, SiII, and CII) that these emission lines are wide and almost rounded. By using the PCS, can be produced the QSO spectra artificially that are useful for investigation of how to discover QSOs and their continuum spectrum classification. By using the weights of the first two PCS can be defined five classes: class Zero and classes from I to IV, these classifications will help to discover the continuum spectrum in the Ly α forest. 21 continuum (Listed in Table 1) have been used upon spectrum of bright QSOs from SDSS with $z \sim 3$ and a signal to noise ratio more than 20 ($S/N > 20$) and apparent magnitude less than 18 ($m_g < 18$). Initially, by investigating the spectrum of each one of these QSOs, the QSOs were classified and the peak of emission line of Ly α belonging to each one of classes was investigated. In the rest, the result mean spectrum of 21 QSOs were compared with mean spectrum of 50 QSOs with redshift $0.14 < z < 1.04$ in Suzuki et al. (2007).

Keywords: PCA statistical method, absorption lines of QSOs, emission lines of QSOs and intergalactic medium

1. Introduction

The importance of investigated absorption lines of QSOs for discovering continuum spectra in the region of the Ly α forest is to study the physical characteristics of intergalactic medium (IGM) and extract the universal parameters. It is impossible to fit the spectrum of the Ly α forest measurements, thus for describing a global plan of continuum, all seeking to find a method for classifying the simple spectra of QSOs (Kirkman et al. 2005, pp. 1373-1380, Suzuki et al. 2003, pp. 1050-1067 and Boroson, 2002, p. 1265). PCA which is known as Karhunen-Loève expansion, is one of the abilities of PCA method is to abbreviate the information existent in a big collection of data and enables one to summarize the information existent in a big collection of data and it is used widely in many zones of astronomy (Hewett et al. 1995, pp. 1498-1521, Croft et al. 1998, p. 44, Bechtold et al. 1994, pp. 1-78 and Kirkman et al. 2003, pp. 1-28). Francis et al. (1992) used PCA method for 232 QSO spectra ($1.8 < z < 2.2$ and wavelength range from 1150 to 2000Å) resulted from Large Bright QSOs (Hewett et al. 2001, pp. 518-535). They showed that the first three PCS involve 75% of variance. Boroson & Green (1992) used from 87 QSO spectra with low redshift ($z < 0.5$) and investigated the relationship between the first two PCS with physical characteristics. Yip et al. (2004) used PCA method for 16707 QSO spectra with $0.08 < z < 5.41$ and wavelength range from 900 to 8000Å and reported that spectrum classification is dependent on redshift and radiance. The continuum spectra of 21 QSOs with $z \sim 3$ were predicted by using PCA method in the region of the Ly α forest, where it is difficult to observe the continuum because of superfluity of absorption lines resulted from IGM. The purpose of this article is to investigate the variety of the spectrum of QSOs according to following in §3, explaining the formulation of PCA method for quantitative description of QSO spectrum using from specific spectra or PCS. In §4, the definition of the theory of QSO artificial spectra and in §5, introduction of five classes of QSO spectra to help qualitative investigation of the variety of QSO spectra.

2. Data

Statistical sample of data was selected out of a collection more than 105000 released QSOs in SDSS, 21 QSOs have redshift $z \sim 3$ and signal to noise ratio more than 20 and apparent magnitude less than 18 (Table 1). By using PCA method, the continuum of these bright QSOs in the region of the Ly α forest was predicted (Suzuki et al. 2005, pp. 592-600). The PCA method, the spectra were transferred to observational spectrum in the rest frame; the wavelength range was defined from 1020 to 1600Å with a step of 0.5Å. To obtain the normal spectrum, the mean flux of 21 pixels obtained around 1280Å, and then the flux of each QSO was divided upon its related mean. On the spectrum resulted from MATLAB software, a norm spectrum was fitted with a condition that this spectrum omits the absorption line and only connects the emission lines to each other. To predict the continuum, it is sufficient to consider the continuum belonging to the red region of spectrum, use this part of the spectrum and predict the complete continuum of QSO in both regions (Kiamehr & Aghae, 2012).

3. PCA and Quantitative Investigation of Rebuilt Spectra

The spectrum of a QSO can be defined in Dirack bracket to the shape of $|q_i\rangle$ which is usually used in quantum mechanics.

$$|q_i\rangle \sim |r_{i,m}\rangle = |\mu\rangle + \sum_{j=1}^m C_{ij}|\varepsilon_j\rangle \quad (1)$$

Where $|q_i\rangle$: is the spectrum of each QSO

$|r_{i,m}\rangle$: is a reconstructed spectrum

$|\mu\rangle$: is the mean of QSOs spectrum

$|\varepsilon_j\rangle$: is the j th PCS

C_{ij} : is weight of each one of PCS. Covariance and correlation matrix of 50 QSO spectrums in Suzuki et al. (2005) were investigated and PCS has been obtained by making a diagonal variance matrix. PCS and their weights would be selected like for becoming orthonormal (Francis et al. 1992, pp. 476-490).

$$\langle \varepsilon_i | \varepsilon_j \rangle = \delta_{ij} \quad (2)$$

$$C_{ij} = \langle q_i - \mu | \varepsilon_j \rangle \quad (3)$$

$$\langle q_i - r_{i,m} | q_i - r_{i,m} \rangle = \sum_{j=m+1}^N C_{ij}^2 \quad (4)$$

Where N is the number of 50 QSO spectra because N is smaller than the number of pixels (1167). The square root of specific values of the j th PCS is defined as:

$$\lambda_j = \frac{1}{N-1} \sqrt{\sum_{i=1}^N C_{ij}^2} \quad (5)$$

$$\sum_{i=1}^N C_{ij} = 0 \quad (6)$$

λ_j for describing the probability distribution function (PDF), the weights of PCS are used (Bechtold, 1993, pp. 143-238, Jannuzi et al. 1996, p. 11, and Croft, 2002, pp. 20-52). The probability distribution function of weight coefficients can only be defined like using a parameter λ_j which has a weight C_{ij} in the distance of $-x_0 \leq C_{ij} \leq x_0$ (Eq. 7) and the spectrum of a QSO can be written as:

$$P(-x_0 \leq C_{ij} \leq x_0) = \int_{-x_0}^{+x_0} \frac{1}{\sqrt{2\pi}\lambda_j} e^{-\frac{x^2}{2\lambda_j^2}} dx \quad (7)$$

$$|q_i\rangle \sim |r_{i,m}\rangle = |\mu\rangle + \sum_{j=1}^m \lambda_j \delta_{ij} |\varepsilon_j\rangle \quad (8)$$

δ_{ij} is the weights of PCS which show the mean deviance of spectrum resulted from principal components of j th

spectrum in comparison with *i*th QSO. The first, the second and the third PCS contain 63.4, 14.5 and 6.2% of variance which means that the first three PCS contain 84.3% of variance (This percentage is dependent on normalization and wavelength range). Share of the variance in the series studied is 84.3% and it is more than what others have claimed (Francis et al. 1992, pp. 476-490 and Shang, 2003, p. 6122).

4. Artificial Spectra

The PCS can be used to create artificial spectrum, artificial spectra can be useful in diagnostic test of QSOs, flow calibration, continuum prediction and cosmos simulator. Artificial spectra can be created by evaluating the weights of PCS. The probability distribution function of weights of *j*th PCS is well showed as a Gaussian function with a zero mean and the standard deviation of λ_j , so if PCS have been added and these weights together, a set of artificial spectra of QSO can be created (Bechtold et al. 2002, pp. 2054-2063 and Cabanac et al. 2002, pp. 1090-1116). The spectrum of QSOs with a flat spectrum and the least absorption lines is mostly used to study the IGM. Therefore, artificial spectra of QSO can be useful in predicting the continuum of QSO in the region of the Ly α forest and for calibrating these continuum spectra on the region of the Ly α forest (Tytler, 2004, pp. 1-28). In figure 1, the predicted continuum according to PCA method for two QSOs out of the collection has been studied.

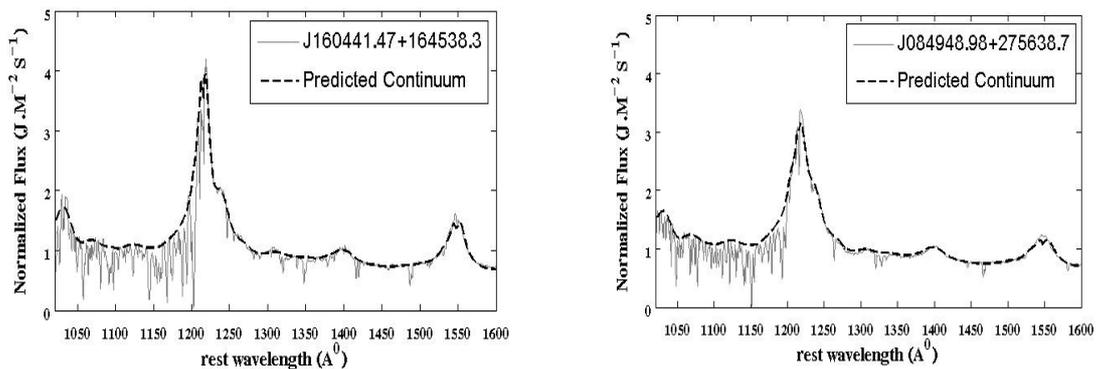


Figure 1. The continuum of the 13th and 4th row QSOs obtained according to PCA method.

5. Classification using PCA method

In this section, by using the definition of the weights of the PCS (meaning the C_{ij}), the QSO spectra classified quantitatively analyzed. The weights of the first two PCS have been used to show the difference between emission lines and continuum spectra and polar specifications are considered like the following:

$$r_i = \sqrt{\delta_{i1}^2 + \delta_{i2}^2} \tag{9}$$

$$\tan \theta_i = \frac{\delta_{i2}}{\delta_{i1}} \tag{10}$$

Where r_i is represents the deviation from the mean spectrum and θ_i is related to the profiles of the emission lines. The diagram δ_{i1} according to δ_{i2} is divided to five regions and five classes of a variety of the QSO spectra are considered. The main purpose of this classification is to differ between similar classes of the QSO spectra (Jena et al. 2004, p. 1552).

Class zero is defined for those which have a limited deviance from the mean spectrum, classes I to IV are explained according to one fourth of diagram δ_{i1} based on δ_{i2} . The probability $r \leq r_0$:

$$P(r \leq r_0) = \int_0^{r_0} e^{-r^2/2} dr = 1 - e^{-r_0^2/2} \tag{11}$$

If $r_0 = 0/668$ then $P(r \leq r_0) = 0/2$. So, class zero is considered for the spectrum of QSOs which has $r \leq 0/668$. For the QSO spectra with $r > 0/668$, classes I to IV are defined which is relevant to the first to the fourth quarter existent in diagram δ_{i1} based on δ_{i2} (Figure 2). The weights of PCS are shown in diagram δ_{i1} based on δ_{i2} in figure 4 that are limited by a circle with a radius of $r_0 = 0/668$ (Suzuki et al. 2007), the number of dots point to the identification of the number of *i* listed QSO in Table 1.

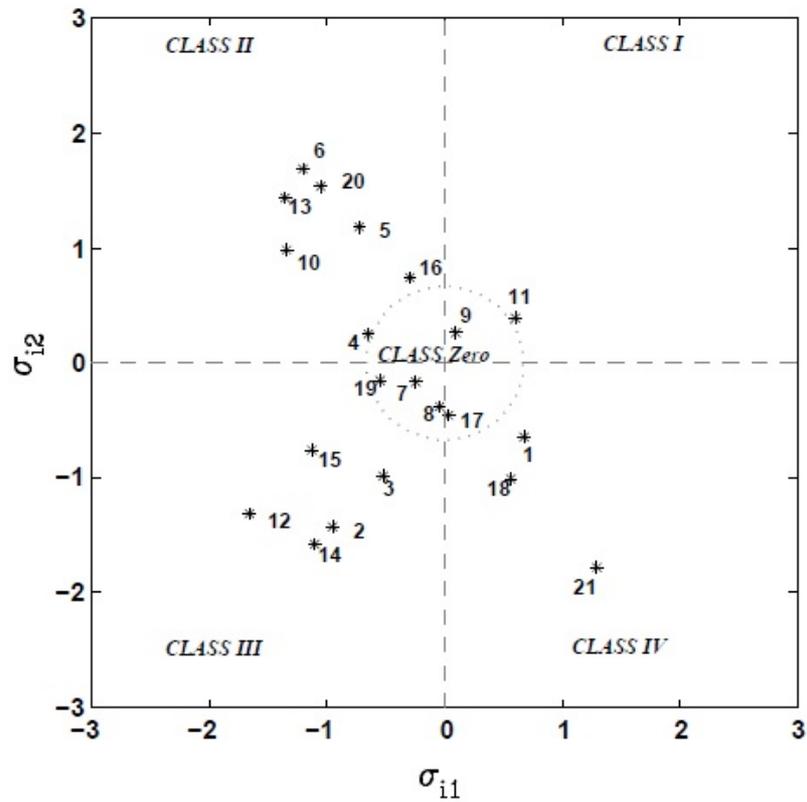


Figure 2. Investigating the standard distribution of weights of the first two PCs for 21 studied QSO, the exact characteristics of each one of the dots is mentioned in Table 1 for all of these QSOs. Five regions for five classes are shown in the figure. Class zero is for defined QSOs which have $r \leq 0/668$ and classes I to IV is for those which have $r > 0/668$

Table 1. The list of QSOs with redshift of $z \sim 3$, S/N more than 20 and m_g less than 18

<i>i</i>	QSO	<i>z</i>	S/N	<i>r</i>		
CLASS Zero						
7	J105236.34+253956.1	2.878	23.6	17.853	0.2994	213.54
8	J093207.46+365745.5	2.893	27	17.696	0.3870	264.06
9	J135044.66+571642.9	2.896	27.3	17.445	0.2839	71.35
17	J090423.37+130920.7	2.968	30.6	17.596	0.4544	265.43
19	J132255.66+391207.9	2.978	25.4	17.752	0.5687	195.16
CLASS I						
11	J152119.68-004818.7	2.9339	25.5	17.932	0.7104	33.01
CLASS II						
4	J084948.98+275638.7	2.848	21.4	17.964	0.7033	158.43
5	J012156.03+144823.9	2.869	31.1	17.422	1.3974	121.46
6	J160843.90+071508.6	2.8775	35.3	17.803	2.0719	125.46
10	J003311.34-171041.5	2.905	20.6	17.973	1.6665	143.79
13	J160441.47+164538.3	2.932	38	16.908	1.9789	133.35
16	J134811.76+281801.8	2.9655	39.4	17.566	0.7994	111.95
20	J120006.25+312630.8	2.984	34	17.619	1.8686	124.57
CLASS III						
2	J020950.71-000506.4	2.841	29.1	17.026	1.7095	236.33
3	J103249.88+054118.3	2.845	28.4	17.347	1.1175	241.69
12	J075326.11+403038.6	2.929	20.9	17.917	2.1111	218.36
14	J111038.63+483115.6	2.953	33	16.792	1.9294	235.18
15	J130337.21+194926.7	2.953	22.5	17.844	1.3600	213.87
CLASS IV						
1	J123549.46+591027.0	2.82	28.9	17.053	0.9312	316.51
18	J120331.29+152254.7	2.9763	32.7	17.933	1.1561	298.82
21	J143912.34+295448.0	2.9923	24.1	17.654	2.2005	305.90

6. The Definition and Explanation for the Characteristics of Five Classes

Figure 3 shows the artificial the QSO spectra produced in the four classes I to IV for explaining the plan of the QSO spectra, they are the result of mean summation of spectrum and the first two PCS with $\delta_{i1} = \pm 1$ and $\delta_{i2} = \pm 1$. It shows four the QSO spectra with the same scales in figure 3 to see the equivalence of emission lines with continuum in a steady mood, in figures 4 to 8, two observed spectra out of each class are shown (With an exception of class I that there is only one QSO in this class). Numbering in figures 4 to 8 is the same as shape 2 to imagine where QSO and its spectrum are in the diagram δ_{i1} according to δ_{i2} .

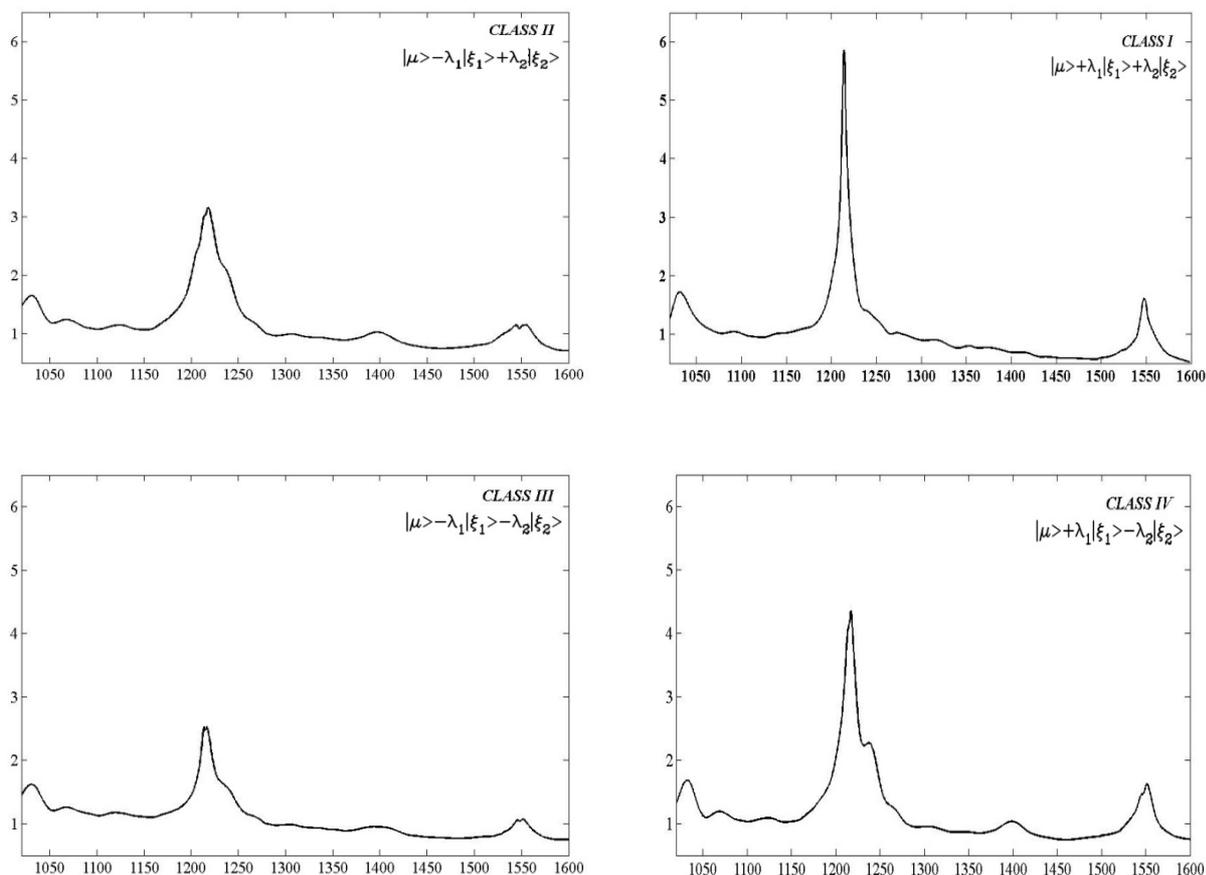


Figure 3. Image making of four classes using the four produced artificial spectra with the first two PCS. Notice that the four spectra are shown in the same horizontal and vertical scales for easier comparisons, so that the vertical axis shows the normalized share and the horizontal axis shows the wavelength. By comparing these spectra, we will see that the spectrum of classes I to IV has more prominent emission lines in comparison with classes II and III

The characteristics of the first two PCS directly show the five classifications, the first PCS has strong and sharp lines of Ly α , Ly β and peaks of emission lines with high ionization (OVI, NV, SiIV, and CIV). The second PCS has peaks of weak emission lines (FeII and FeIII in the blue part of the Ly α emission line and SiII, and CII in the red part of the spectrum) and their vertical slices are round and wide. In the second PCS, the size of emission lines share with low ionization and strong the Ly α emission lines and CIV have opposite sides which mean that they are not dependent together and they are not associated. Totally, emission lines of Ly α and CIV explain un-concurrent vertical slices for these emission lines. According to the definition, the first two PCS guarantee the correlation between emission lines, for example, if a QSO shows a peak with low ionization in the red part of the Ly α emission line (SiII, and CII), it must have the second weight of principal component of negative spectrum and must expect to see it prominent in the region of the Ly α forest of FeII and FeIII, so, such a QSO must belong to one of the classes III or IV. If other QSOs have Ly α , Ly β and emission line with strong and sharp high ionization (NV, SiIV, and CIV), they must have the first positive weight of PCS and belong to classes I or IV. The emission line of Ly α and NV resulted from normalized share have the biggest peak in class I among five classes. The classes I and II or the classes III and IV can be defined via the combination of said characteristics. In reality, diagonal classes have opposite characteristics, for example, class I has emission lines with sharp and high ionization where class III has wide emission lines.

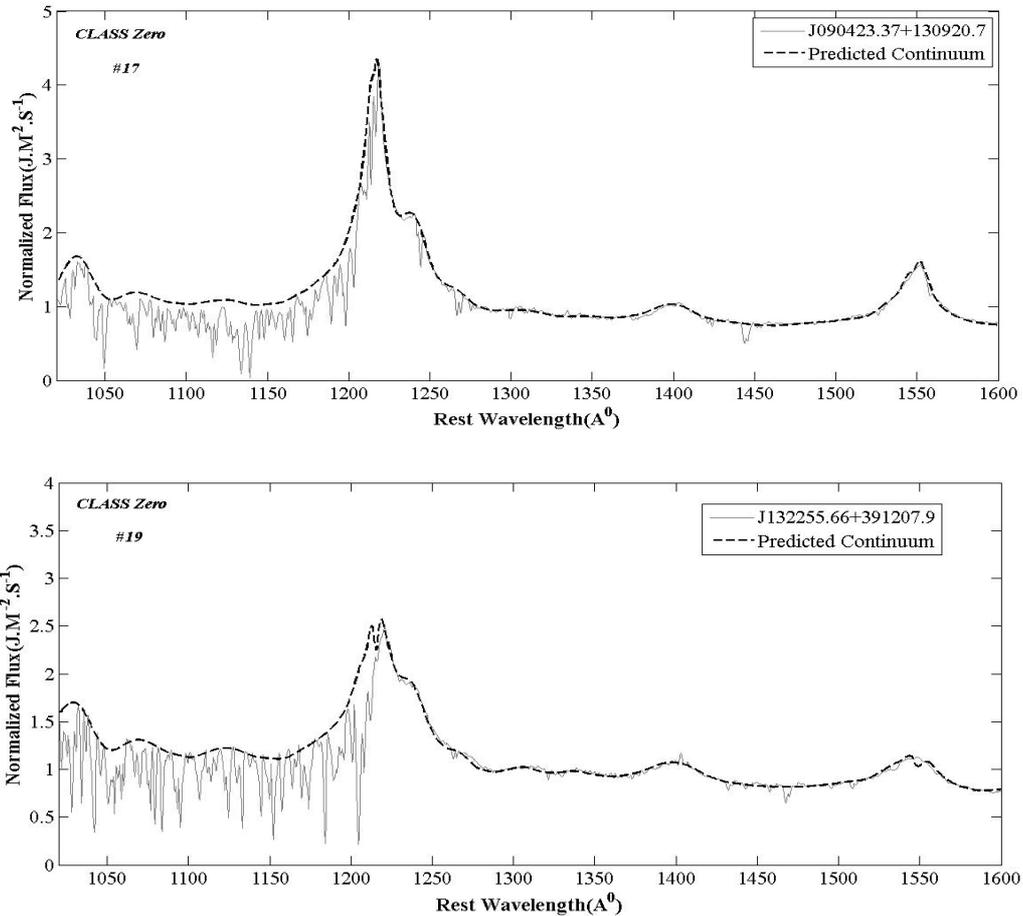


Figure 4. Class zero ($r \leq 0.668$) the diagram of plain line shows the observed spectrum and the diagram of dotted line shows the continuum resulted from PCA method

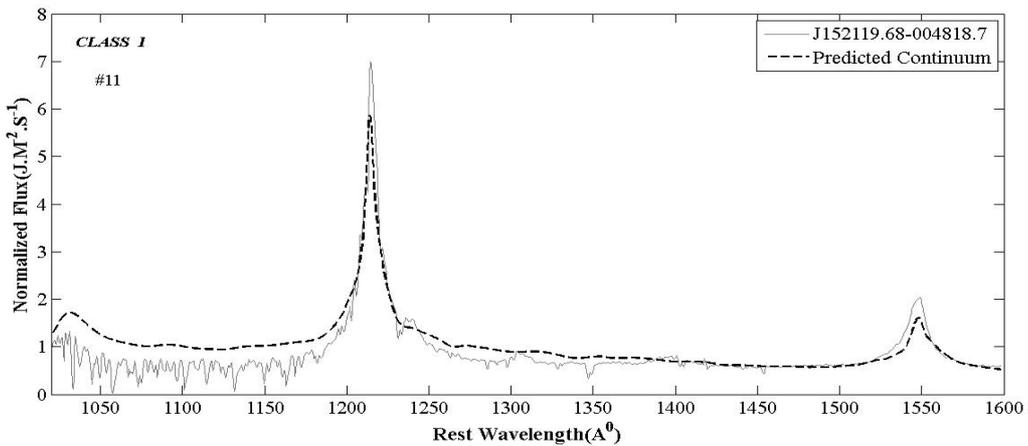


Figure 5. Class I ($r > 0.668, 0 \leq \theta \leq 90$) the diagram of plain line shows the observed spectrum and the diagram of dotted line shows the continuum resulted from PCA method

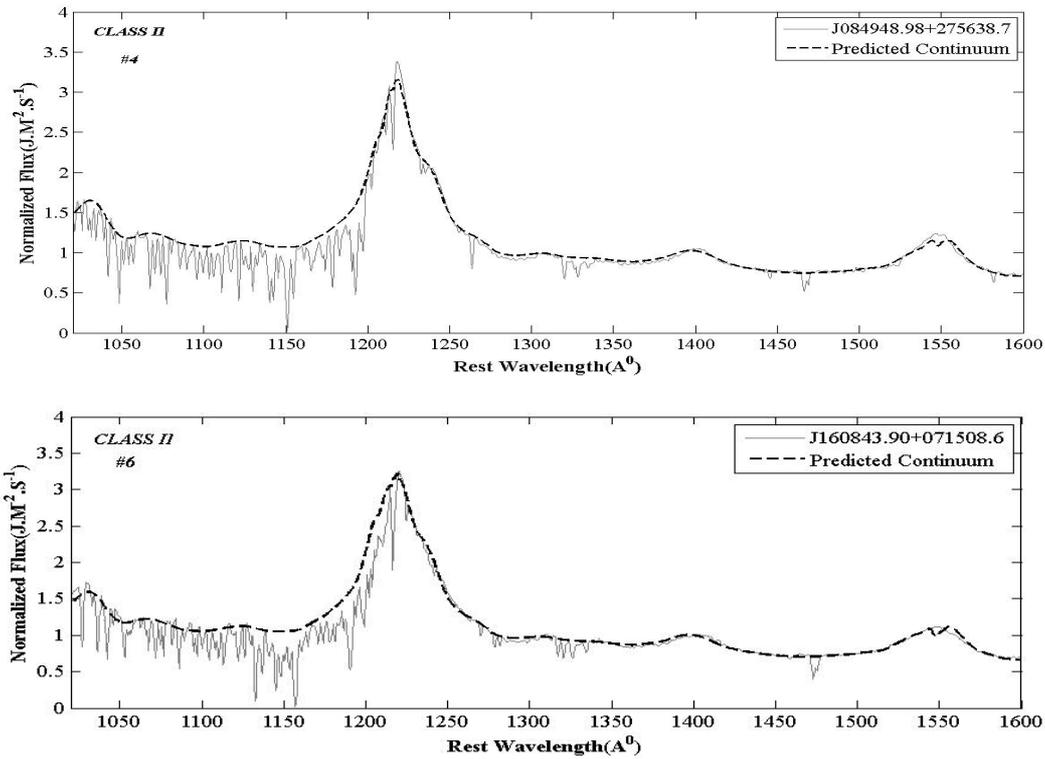


Figure 6. Class II ($r > 0.668, 90 \leq \theta \leq 180$) the diagram of plain line shows the observed spectrum and the diagram of dotted line shows the continuum resulted from PCA method

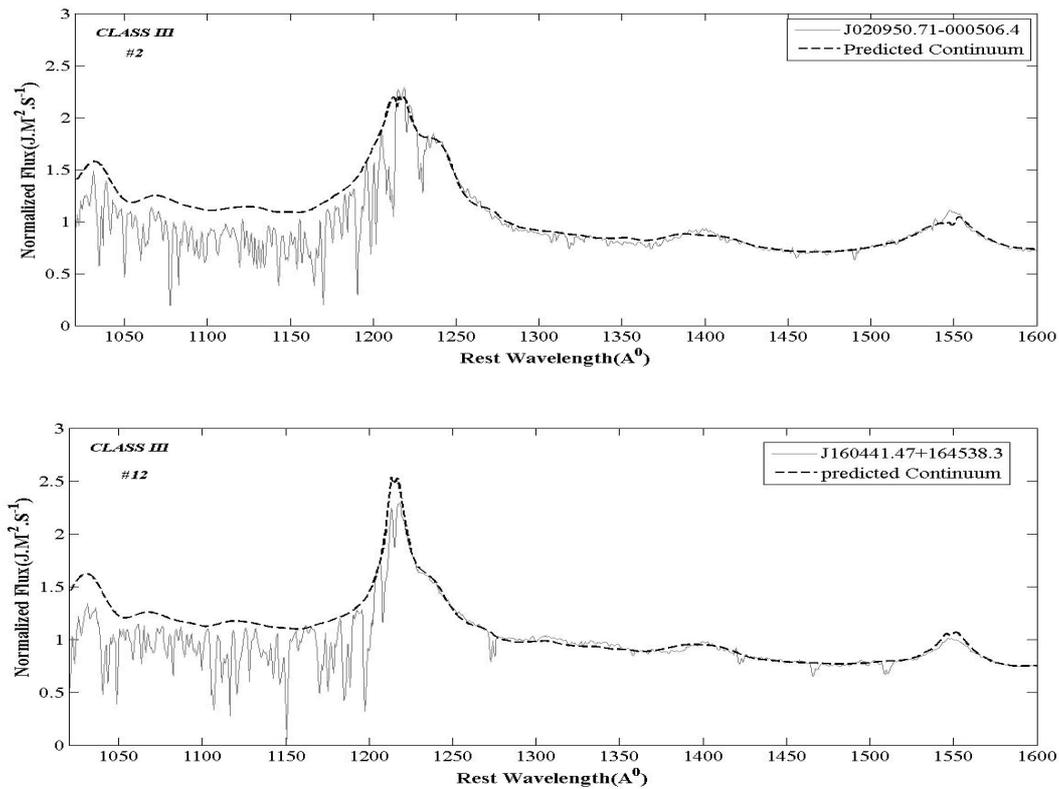


Figure 7. Class III ($r > 0.668, 180 \leq \theta \leq 270$) the diagram of plain line shows the observed spectrum and the diagram of dotted line shows the continuum resulted from PCA method

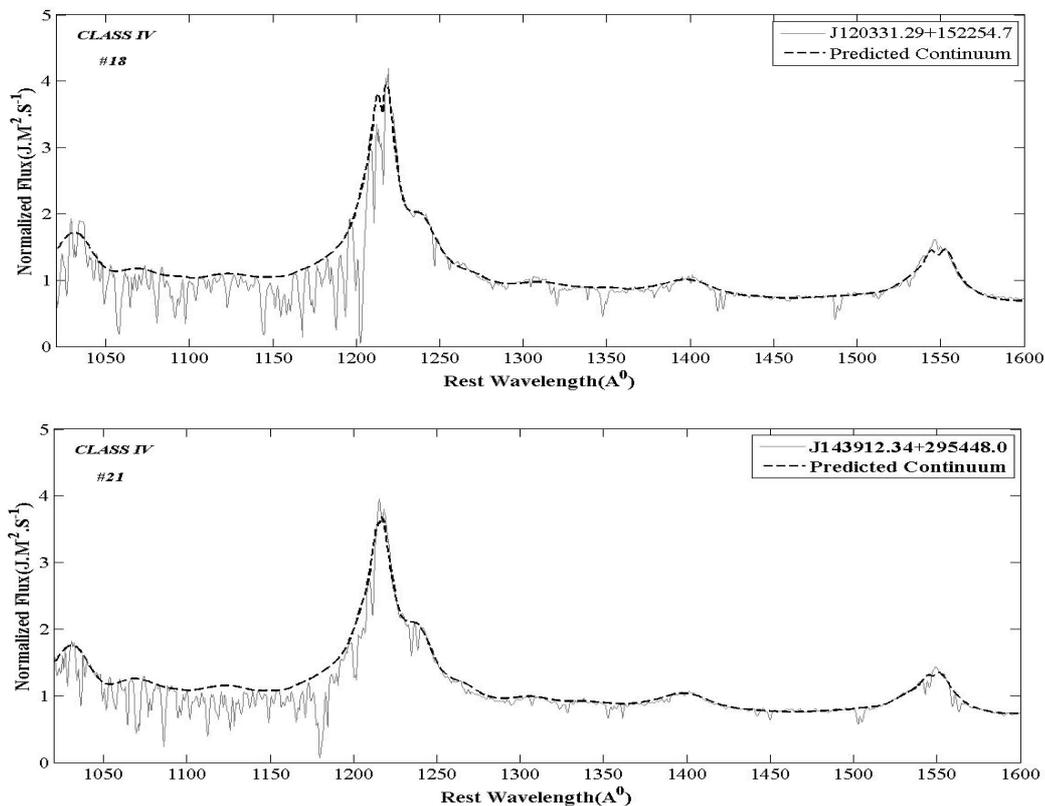


Figure 8. Class IV ($r > 0.668$, $270 \leq \theta \leq 360$) the diagram of plain line shows the observed spectrum and the diagram of dotted line shows the continuum resulted from PCA method

7. Conclusion

A wide variety of vertical slices of emission lines have been in both analyzed and investigated in the region of the Ly α forest qualitative and quantitative moods. The variety of QSO spectra has been described using PCA method (PCA) and understood that 1161 pixels (in the wavelength range from 1020 to 1600Å with a step of 0.5Å) can be summarized using the weights of the first seven PCS because pixels are not independent but they are so correlated and dependent to each other. In figure 9, the diagram related to the mean of 50 continuums of QSOs with $z < 1$ and dotted line diagram related to the mean of 21 continuums of QSOs with $z \sim 3$ are shown, the main difference between two diagrams is about the peaks of their emission lines. The reason of this difference can be explained by paying attention to figure 2, by comparing figure 2 of this article and figure 4 of Suzuki et al. (2007), there are a few QSOs in classes I to IV with high peaks here, whereas in Suzuki et al. (2007), most of the QSOs are located in these two classes. So, the peak of emission lines of linear diagram of the mean spectrum of 50 QSOs must be higher than the peak of emission lines of dotted line diagram of the mean spectrum of 21 QSOs here and it can be well seen in the figure. In Suzuki et al. (2007), the theory of creating artificial spectra of QSO must be useful in identifying, calibrating and simulating. The five classes have been introduced to segregate similar spectra of QSOs and show that how the classification of spectra can guide to find continuum in the region of the Ly α forest.

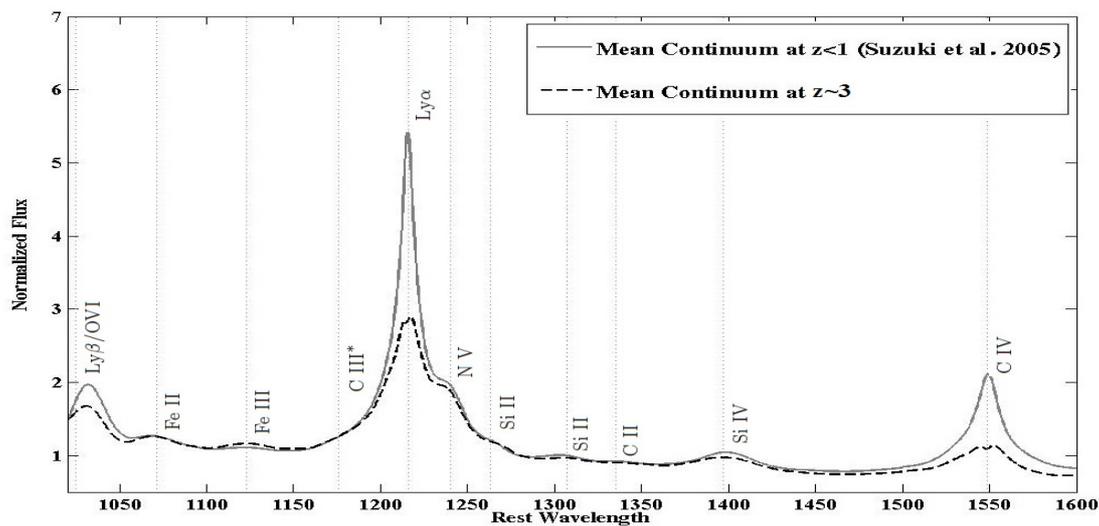


Figure 9. Linear diagram is related to the mean of 50 continuums of QSOs with redshift $z < 1$ and dotted line diagram is related to the mean of 21 continuums of QSOs with redshift $z \sim 3$

References

- Bechtold, J. (1994). The Lyman-Alpha forest near 34 quasi-stellar objects with Z greater than 2.6. *Astrophysical Journal Supplement Series (ISSN 0067-0049)*, 91, 1-78.
- Bechtold, J., Aldcroft, L., & Elvis, M. (2002). A sample of lobe-dominated quasars with Mg II absorption lines. *Astronomical Journal (ISSN 0004-6256)*, 105, 2054-2063.
- Bechtold, J., et al. (1993). A Uniform Analysis of the Ly α Forest at $z = 0-5$. III. Hubble Space Telescope Faint Object Spectrograph Spectral Atlas. *The Astrophysical Journal Supplement Series*, 140, 143-238.
- Boroson, A. (2002). What Causes the Scatter in the Black Hole Mass vs. Velocity Dispersion Relation for AGNs?, American Astronomical Society, 201st AAS Meeting, #98.01. *Bulletin of the American Astronomical Society*, 34, 1265.
- Boroson, A., & Green, F. (1992). The emission-line properties of low-redshift quasi-stellar objects. *Astrophysical Journal Supplement Series (ISSN 0067-0049)*, 80, 109-135.
- Cabanac, A., de Lapparent, V., & Hickson, P. (2002). Classification and redshift estimation by principal component analysis. *Astronomy and Astrophysics*, 389, 1090-1116.
- Croft, C., et al. (2002). Toward a Precise Measurement of Matter Clustering: Ly α Forest Data at Redshifts 2-4. *The Astrophysical Journal*, 581, 20-52.
- Croft, C., Weinberg, H., Katz, N., & Hernquist, L. (1998). Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Ly alpha Forest (p. 44). *Astrophysical Journal*, 495.
- Francis, J., Hewett, C., Foltz, B., & Chaffee, H. (1992). An objective classification scheme for QSO spectra (pp. 476-490). *Astrophysical Journal, Part 1 (ISSN 0004-637X)*, 398.
- Hewett, C., Foltz, B., & Chaffee, H. (1995). The large bright quasar survey 6: Quasar catalog and survey parameters (pp. 1498-1521). *Astronomical Journal (ISSN 0004-6256)*, 109.
- Hewett, C., Foltz, B., & Chaffee, H. (2001). The Large Bright Quasar Survey. VII. *The LBQS and FIRST Surveys. The Astronomical Journal*, 122, 518-535.
- Jannuzi, T., et al. (1996). The Hubble Space Telescope Quasar Absorption Line Key Project: The Unusual Absorption-Line System in the Spectrum of PG 2302+029---Ejected or Intervening? *Astrophysical Journal Letters*, 470, 11.
- Jena, T., Norman, L., Tytler, D., & Suzuki, N. (2004). Precise Measurement of the Matter Power Spectrum Amplitude and the Background Radiation Amplitude, American Astronomical Society Meeting 205, #123.06. *Bulletin of the American Astronomical Society*, 36, 1552.

- Kiamehr, Z., & Aghaee, A. (2012). *Determination of the quasar continuum using the principal components of spectrum*, M.Sc. Thesis, university of Sistan and Baluchestan, Zahedan.
- Kirkman, D., et al. (2005). The HI opacity of the intergalactic medium at redshifts $1.6 < z < 3.2$. *Monthly Notices of the Royal Astronomical Society*, 360, 1373-1380.
- Kirkman, et al. (2003). The Cosmological Baryon Density from the Deuterium-to-Hydrogen Ratio in QSO Absorption Systems: D/H toward Q1243+3047. *The Astrophysical Journal Supplement Series*, 149, 1-28.
- Shang, Z. (2003). Study of ultraviolet-optical properties of a complete sample of QSOs, ProQuest Dissertations and Theses; Thesis (Ph.D.)--The University of Texas at Austin, Publication Number: AAT3116394; ISBN: 9780496638291; Source: Dissertation Abstracts International, 64-12, 6122.
- Suzuki, N. (2007). Quasar Spectrum Classification with PCA - II: Introduction of Five Classes, Artificial Quasar Spectrum, the Mean Flux Correction Factor dF , and the Identification of Emission Lines in the Ly α Forest, eprint arXiv:astro-ph/0503248.
- Suzuki, N., Tytler, D., Kirkman, D., O'Meara, M., & Lubin, D. (2003). Relative Flux Calibration of Keck HIRES Echelle Spectra. *The Publications of the Astronomical Society of the Pacific*, 115, 1050-1067.
- Suzuki, N., Tytler, D., Kirkman, O., John, M., & Lubin, D. (2005). Predicting QSO Continua in the Ly α Forest. *Astrophysical Journal*, 618, 592-600.
- Tytler, D., et al. (2004). Cosmological Parameters σ_8 , the Baryon Density Ω_b , the Vacuum Energy Density Ω_Λ , the Hubble Constant and the UV Background Intensity from a Calibrated Measurement of HI Ly α Absorption at $z=1.9$ (pp. 1-28). *The Astrophysical Journal*, 617.
- Tytler, D., et al. (2004). Mean amount of Absorption from the Intergalactic Medium, American Astronomical Society Meeting 205, #129.06. *Bulletin of the American Astronomical Society*, 36, 1557.
- Yip, W., et al. (2004). Spectral Classification of Quasars in the Sloan Digital Sky Survey: Eigenspectra, Redshift, and Luminosity Effects (pp. 2603-2630). *The Astronomical Journal*, 128.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).