EasyChair Preprint

№ 7766

Principal Component Analysis and Factor Analysis to Analyze the Different Arrangements About the Quran's Suras

Yanwen Wang, Javad Garjami, Milena Tsvetkova,
Nguyen Huu Hau, Kim-Hung Pho and
Mohammad Reza Mahmoudi

April 12, 2022

Principal Component Analysis and Factor Analysis to Analyze the Different Arrangements about the Quran's Suras

Yanwen Wang

School of Literature, Shandong University, Jinan, Shandong Province, China

Javad Garjami

Arabic Language and Literature, Faculty of Literature and Humanities, Mohaghegh Ardabili University, Ardabili, Iran

Milena Tsvetkova

Faculty of Journalism and Mass Communication, Sofia University, Sofia, Bulgaria

Nguyen Huu Hau

Department of Training Management, Hong Duc University, Thanh Hoa City, Vietnam

Kim-Hung Pho

Fractional Calculus, Optimization and Algebra Research Group, Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam

Mohammad Mahmoudi

Fractional Calculus, Optimization and Algebra Research Group, Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam

Abstract. Data mining, statistics and data analysis are popular techniques to study datasets and extract knowledge from them. In this paper, principal component analysis and factor analysis were applied to cluster thirteen different given arrangements about the Suras of the Holly Quran. The results showed that these thirteen arrangements can be categorized in two parts such that the first part includes Blachère, Davood, Grimm, Nöldeke, Bazargan, E'temad-al-Saltane and Muir, and the second part includes Ebn Nadim, Jaber, Ebn Abbas, Hazrat Ali, Khazan and Al-Azhar.

Keywords: Quran, Principal Component Analysis, Factor Analysis, Statistics, Computer Science.

1. Introduction

The Quran is the most sacred, most ancient, and most important book of Muslims. After the Gospel and the Torah, it was revealed to the Prophet by God through Gabriel. The main purpose of the Quran, which is pursued from beginning to end, is to know the essence of God and to guide man. But there are various subjects such as emotions, ethics, beliefs, personal and social upbringing and people are called to be a complete divine personality. The Muslims believes that the Holy Quran was revealed to Mohammed as the prophet of God through Gabriel for twenty three years in Mecca and Medina, under various conditions at any time of day and night. They believe that the all of the words given by Gabriel are the God's words. It may be called by other names such as Moshaf or Forqan. The Holy Quran includes 114 Suras. Each Sura chapter contains several verses with a beginning and an ending. There are more than 6000 verses in Quran.

The revelation of the verses of the Quran from Gabriel to Mohammed continued to the last years of his life. Therefore, it was impossible to finalize a copy of it based on the times of the revelations. In other words, although all of the Quran's Suras were written during the time, but the order or sequence of the Suras and verses were not arranged [Suyuti (1998)]. Nevertheless,

many of the Muslims believed that the arrangement of the Suras was not based on the exact revelation and the order of Suras from Hamd to Nas was originally different from the true order of revelation. The expert scientists believe that the meaning of the Quran in line with its revelation is very important. Because the reader can simpler understand the bavkground of the revelation, the true meanings, and their reasons [Eskandarlu (2008)]. Accordingly, there have been numerous attempts throughout history to find a valid timetable that can reveal the true order of revelation. For example, Ibn Nadi, Jaber ibn Zayd, and many others paid attention to the ordering of the Suras, in addition to many Orientalists such as Hirschfeld (1902), Blachere (1947), Bell (1953), and Noldeke (2012) have considered this purpose.

As it can be seen in Hirschfeld (1902), Weil considered the historical and the compilation order of the verses and the Suras. He divided the Suras based on their subjects and texts. Blachere (1947) and Bell (1953) investigated the Quran and its translation and published several works in this regard. Noldeke (2012) studied and discussed about the several compilations of the Quran.

In this study, some statistical methods such as principal component analysis and factor analysis are applied to cluster different given arrangements about the Suras of the Holly Quran.

2. Literature Review

Naji et al. (2005) designed a classifier to categorize the verses in each Sura. Bin Dost and Ahmad (2008) investigated the structural properties of the Meccan and the Medinan Surras. They studied the descriptive statistics for the size and the length of the words of the Surras. Abdul-Baquee and Atwell (2009) examined the verbs in the Quran and compared it to the frames and verbs in the English frame grid. Sadeghi (2011) endorsed a timeline in which seven groups of crossings represent successive stages. Alhawarat et al. (2015) began a series of researches aimed at

providing accurate and useful and knowledge and information from the Quran. Moreover, by using statistical methods, they constructed a framework for the scientists in the area of the Arabic natural language processing. Data mining and analysis are popular techniques to study datasets and extract knowledge from them. Mahmoudi et al. (2018a) applied clustering analysis and chi-square test to study "Love" and its alternatives in Divan of Moulana. Mahmoudi and Abbasalizadeh (2018a) applied clustering analysis and chi-square test to study "Love" and its alternatives in poems of Saadi. Mahmoudi and Abbasalizadeh (2018b) applied statistical techniques to study the Divan of Khaghani. Mahmoudi and Abbasalizadeh (2018c) considered the similarities between different orders about the revelation of Quran, based on the hierarchical clustering method and the regression analysis. Yin et al. (2019) used text mining techniques to investigate the Divan of Khaghani. Liu et al. (2019) applied statistical techniques to study the different traits of God in the Meccan and Medinan suras of Quran. Many other scientists applied data mining and analysis in different fields [see for example, Haghbin et al. (2011); Mahmoudi and Mahmoudi [(2014a), (2014b)], Mahmoudi et al. (2016), Mahmoudi et al. [(2017a), (2017b)], Jalali et al. (2017), Maleki and Mahmoudi (2017), Maleki et al. (2017), Bahrami et al. (2017), Mahmoudi (2018), Jalali et al. (2018), Mahmoudi et al. [(2018b), (2018c), (2018d), (2018e)], Heydari et al. (2018), Abbasi et al. (2018), Liu et al. (2019), Yin et al. (2019), Maleki et al. (2019), Ji-jun et al. (2019), Mahmoudi et al. (2019)].

3. Materials and Methods

3.1. Materials

With regard to the fact that the revelation of the verses of the Quran from Gabriel to Mohammed continued to the last years of his life, it was impossible to finalize a copy of it based on the times of the revelations. Therefore, there are different arrangements about the Quran's Suras. In this

research, thirteen famous arrangements about the Quran's Suras are considered. The names of these arrangements are given in Table 1. Each of them arranges the Suras from 1 to 114.

Table 1: Different arrangements about the Quran's Suras

	Name in Persian	Name in English	Label
1	ابن نديم	Ebn Nadim	A_1
2	بلاشر	Blachère	A_2
3	داوود	Davood	A_3
4	گريم	Grimm	A_4
5	نولدكه	Nöldeke	A_5
6	جابر	Jaber	A_6
7	بازرگان	Bazargan	A_7
8	ابن عباس	Ebn Abbas	A_8
9	حضرت علي	Hazrat Ali	A_9
10	اعتمادالسلطنه	E'temad-al-Saltane	A_{10}
11	موير	Muir	A_{11}
12	خازن	Khazan	A_{12}
13	الازهر	Al-Azhar	A_{13}

3.2. Statistical Methods

Thirteen famous arrangements about the Quran's Suras were fed into the computer item by item according to their own values, and were analyzed using the Statistical R software (version 3.6.1). Then, the principal component analysis and the factor analysis were applied to cluster these different given arrangements about the Suras of the Holly Quran.

3.2.1. Principal Component Analysis

Principal component analysis (PCA, in abbreviation) is a famous multivariate technique that converts several correlated variables into several linearly uncorrelated variables named principal components. In this conversion, the first principal components contain the most information about the dataset [Johnson and Wichem (2002)]. In applications, PCA is applied to transform a high-dimensional dataset to a lower-dimensional dataset, by using only the first few principal components so that the dimensionality of the transformed data is reduced. Based on the Kaiser Index, the number of the important principal components is considered as the number of the eigen-values of the correlation matrix that are more than 1.

3.2.2. Factor Analysis

Factor analysis (FA, in abbreviation) similar to PCA, is another famous multivariate technique that converts several correlated variables into other variables named factors. In this conversion, the first factors contain the most information about the dataset [Johnson and Wichem (2002)]. Different from PCA, FA focuses on the correlations of variables such that the variables in a factor are highly correlated with each other and the variables in different factors are highly uncorrelated with each other. Similar to PCA, in applications, FA is applied to transform a high-dimensional dataset to a lower-dimensional dataset, by using only the first few factors so that the dimensionality of the transformed data is reduced. As PCA, Based on the Kaiser Index, the number of the important factors in FA is considered as the number of the eigen-values of the correlation matrix that are more than 1.

4. Results

In this section, the results of PCA and FA about the considered arrangements are reported. Subsection 4.1 is devoted to the results of PCA. The results of FA are also given in Subsection 4.2.

4.1. Results of PCA to Study the Different Arrangements of the Quran's Suras

In this part, the results of PCA to categorize different arrangements for Quran's Suras are reported. Figure 1 shows the eigen-values of the correlation matrix of the variables. As it can be seen, just the first two values are more than 1. Therefore, based on the Kaiser Index, the number of the important components in PCA is considered as 2. Table 2 shows that the first and the second components can determine about 74.589% and 12.551% of the variability of the observations. In other words, we can reduce the dimensions of the variables from 13 to 2. In this case, about 87.140% of the variability of the observations is determined that is an acceptable percent. As Figure 2 indicates, these thirteen arrangements can be categorized in two parts such that the first part includes Blachère, Davood, Grimm, Nöldeke, Bazargan, E'temad-al-Saltane and Muir, and the second part includes Ebn Nadim, Jaber, Ebn Abbas, Hazrat Ali, Khazan and Al-Azhar.

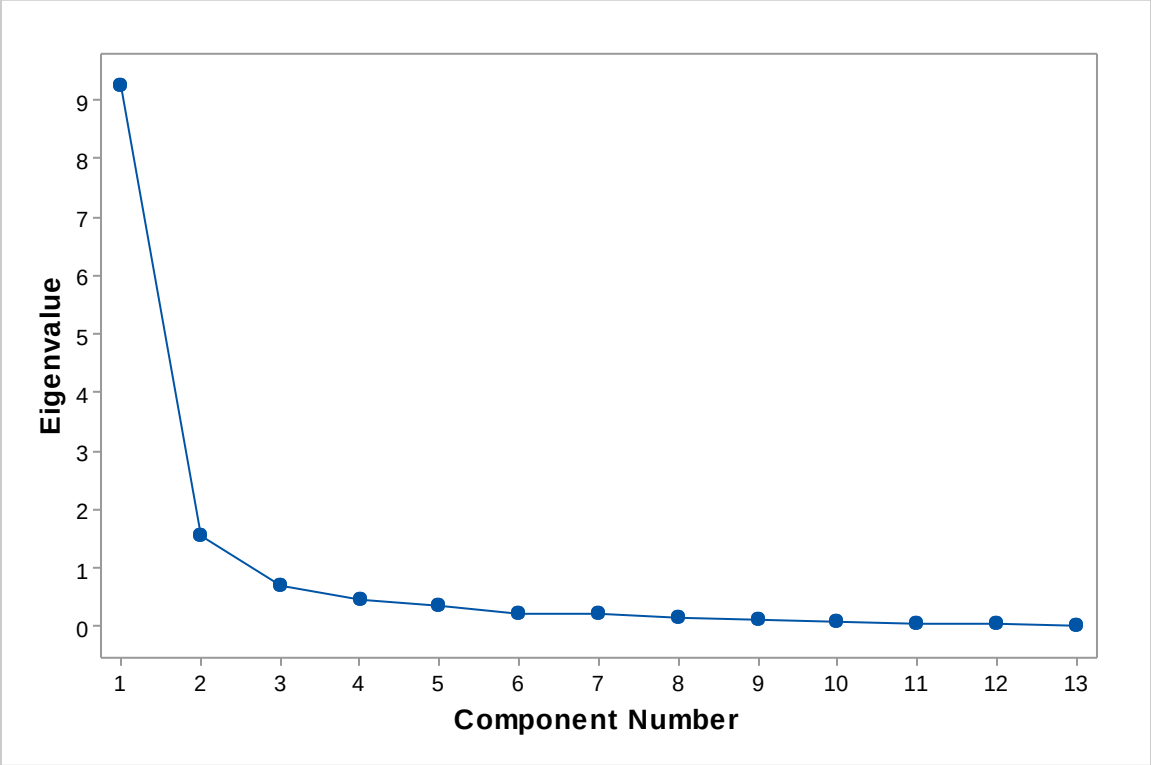

Figure 1: The Scree plot for the eigen-values of the correlation matrix of the variables

Table 2: Total variance explained by each component

Component	Eigenvalues		
	Total	% of Variance	Cumulative %
1	9.697	74.589	74.589
2	1.632	12.551	87.140
3	0.515	3.960	91.100
4	0.358	2.753	93.853
5	0.262	2.016	95.870
6	0.163	1.252	97.121
7	0.132	1.019	98.140
8	0.100	0.767	98.907
9	0.060	0.462	99.369
10	0.042	0.325	99.694
11	0.033	0.255	99.949
12	0.006	0.047	99.996
13	0.001	0.004	100.000

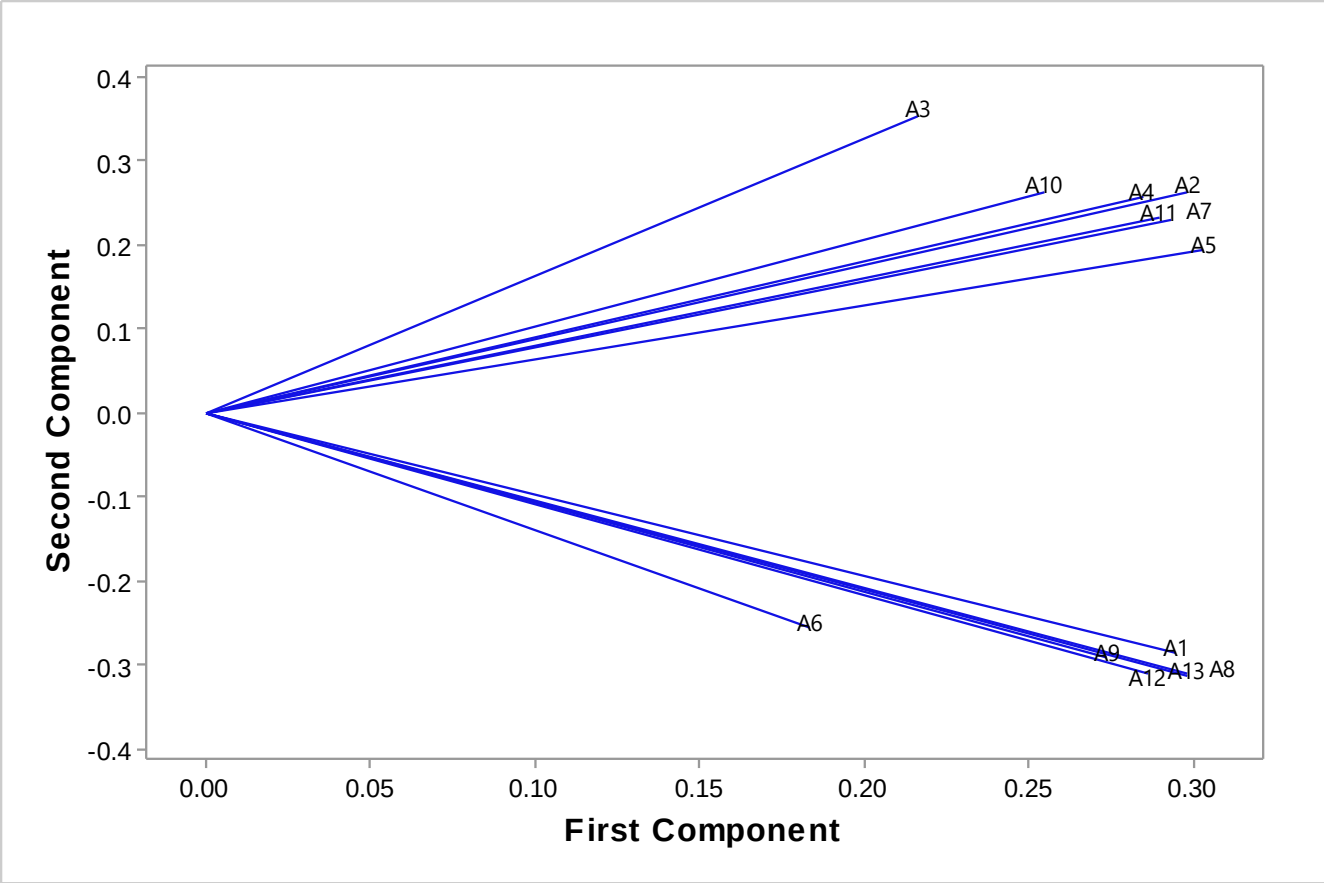

Figure 2: Loading components for the first two components

4.2. Results of FA to Study the Different Arrangements of the Quran's Suras

In this part, the results of FA to categorize different arrangements for Quran's Suras are reported. As PCA, based on the Kaiser Index, the number of the important factors in FA is considered as 2. Table 2 shows that the first and the second components can determine about 74.589% and 12.551% of the variability of the observations. In other words, we can reduce the dimensions of the variables from 13 to 2. In this case, about 87.140% of the variability of the observations is determined that is an acceptable percent. As Figure 3 indicates, these thirteen arrangements can be categorized in two parts such that the first part includes Blachère, Davood, Grimm, Nöldeke,

Bazargan, E'temad-al-Saltane and Muir, and the second part includes Ebn Nadim, Jaber, Ebn Abbas, Hazrat Ali, Khazan and Al-Azhar.

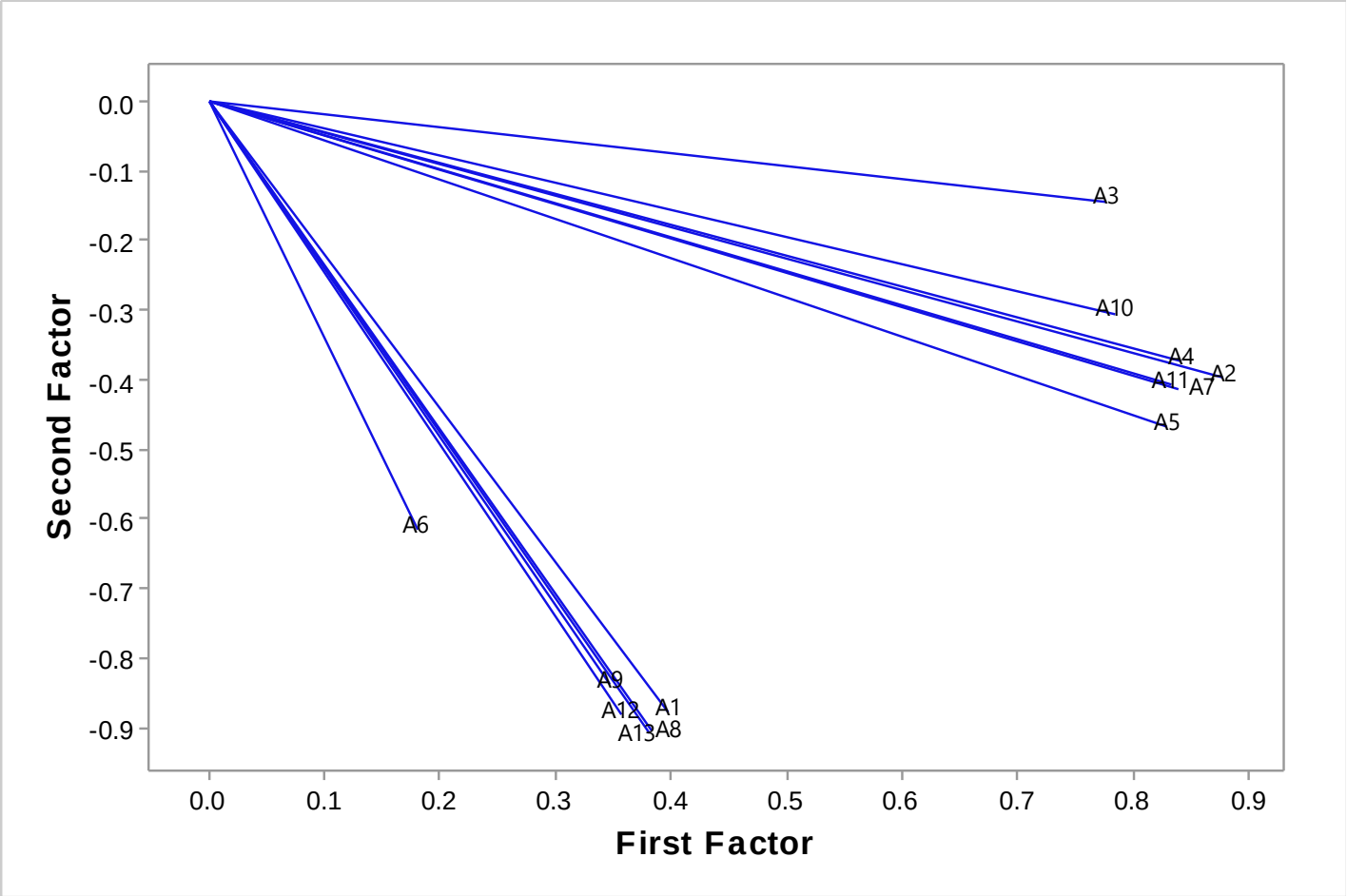

Figure 3: Loading components for the first two factors

5. Discussion

The Quran is the most sacred, most ancient, and most important book of Muslims. After the Gospel and the Torah, it was revealed to the Prophet by God through Gabriel. The Muslims believes that the Holy Quran was revealed to Mohammed as the prophet of God through Gabriel for twenty three years in Mecca and Medina, under various conditions at any time of day and

night. The revelation of the Quran from Gabriel to Mohammed continued to the last years of his life. Therefore, it was impossible to finalize a copy of it based on the times of the revelations. In other words, although all of the Quran's Suras were written during the time, but the order or sequence of the Suras and verses were not arranged. Nevertheless, many of the Muslims believed that the arrangement of the Suras was not based on the exact revelation and the order of Suras from Hamd to Nas was originally different from the true order of revelation. The expert scientists believe that the meaning of the Quran in line with its revelation is very important. Because the reader can simpler understand the bavkground of the revelation, the true meanings, and their reasons. Accordingly, there have been numerous attempts throughout history to find a valid timetable that can reveal the true order of revelation. In this study, some statistical methods such as PCA and FA were applied to cluster thirteen different given arrangements about the Suras of the Holly Quran. The results indicated that these thirteen arrangements can be categorized in two parts such that the first part includes Blachère, Davood, Grimm, Nöldeke, Bazargan, E'temad-al-Saltane and Muir, and the second part includes Ebn Nadim, Jaber, Ebn Abbas, Hazrat Ali, Khazan and Al-Azhar.

References

Abbasi, A. R., Mahmoudi, M. R., Avazzadeh, Z. (2018). Diagnosis and clustering of power transformer winding fault types by cross-correlation and clustering analysis of FRA results, *IET Generation, Transmission & Distribution* **12 (19)**: 4301-4309.

Abdul-Baquee, S., Atwell, E. S. (2009). Knowledge representation of the Quran through frame semantics: a corpus-based approach, in*: Proceedings of the Fifth Corpus. Linguistics Conference*.

Alhawarat, M., Hegazi, M., Hilal, A. (2015). Processing the Text of the Holy Quran: a Text Mining *(IJACSA) International Journal of Advanced Computer Science and Applications* **6(2)**: 262-267.

Bahrami, M., Amiri, M. J., Mahmoudi, M. R., Koochaki, S. (2017). Modeling caffeine adsorption by multi-walled carbon nanotubes using multiple polynomial regression with interaction effects, *Journal of water and health* **15 (4):** 526-535.

Bell, R., (1953). *Introduction to the Qurʾān*. Islamic Surveys, Edinburgh University Press.

Blachère, R., (1947). *Introduction to the Koran,* publisher: the University of California.

Bin Dost, M. Kh., Ahmad, M. (2008). Statistical Profile of Holy Quran and Symmetry of Makki and Madni Surras, *Pakistan Journal of Commerce and Social Sciences* **1**: 1-16.

Eskandarlu, M, J. (2008). *Orientalists and the dating of Quran*, 1st edt., Interpretation researches and Quranic sciences, Qom.

Haghbin, H., Mahmoudi, M. R., Shishebor, Z. (2011). Large Sample Inference on the Ratio of Two Independent Binomial Proportions. Journal of Mathematical Extension **5(1):** 87- 95.

Heydari, M. H., Mahmoudi, M. R., Shakiba, A., Avazzadeh, Z. (2018). Chebyshev cardinal wavelets and their application in solving nonlinear stochastic differential equations with fractional Brownian motion, *Communications in Nonlinear Science and Numerical Simulation* **64:** 98-121.

Jafar Jalali, S.M., Mahdizadeh, E., Mahmoudi, M.R., Moro, S. (2018). Analytical assessment process of e-learning domain research between 1980 and 2014, *Int. J. Management in Education* **12(1)**: 43–56.

Jalali, S. M., Moro, S., Mahmoudi, M. R., Ghaffary, K. A., Maleki, M., Alidoostan, A. (2017). A comparative analysis of classifiers in cancer prediction using multiple data mining techniques. *International Journal of Business Intelligence and Systems Engineering* **1 (2),** 166-178.

Ji-jun, P., Mahmoudi, M. R., Baleanu, D., Maleki, M. (2019). On Comparing and Classifying Several Independent Linear and Non-Linear Regression Models with Symmetric Errors. *Symmetry* **11(6):** 820.

Johnson, R. A., Wichern, D. (2002). *Multivariate Analysis*, John Wiley & Sons, Ltd.

Hirschfeld, H. (1902). *New Researches into the Composition and Exegesis of the Qoran*, Royal Asiatic society, London.

Liu, J., Mahmoudi, M. R., Abbasalizadeh, A. (2019). Statistical analysis about the God's traits in Quran. *Digital Scholarship in the Humanities*. In Press.

Mahmoudi, M. R. (2018). On Comparing Two Dependent Linear and Nonlinear Regression Models. *Journal of Testing and Evaluation.* **47 (1).** DOI: 10.1520/JTE20170461, In Press.

Mahmoudi, M. R., Abbasalizadeh, A. (2018a) 'On comparing and clustering the alternatives of love in Saadi's lyric poems (Ghazals)', *Digital Scholarship in the Humanities. DOI:10.1093/llc/fqy024. In Press.*

Mahmoudi, M. R., Abbasalizadeh, A. (2018b) 'Statistical analysis about the order of Quran's revelation', *Digital Scholarship in the Humanities. DOI:10.1093/llc/fqy030. In Press.*

Mahmoudi, M. R., Abbasalizadeh, A. (2018c) 'How Statistics and Text Mining Can be Applied to Literary Studies?', *Digital Scholarship in the Humanities.* In Press.

Mahmoudi, M. R., Abbasalizadeh, A. and Rahmati, M. (2018a) *'An Statistical Approach to Investigate the Alternatives of Love in Moulana's Divan', International Journal of Business Intelligence and Data Mining*, In Press.

Mahmoudi, M. R., Behboodian, J., Maleki, M. (2017a). Large Sample Inference about the Ratio of Means in Two Independent Populations, *Journal of Statistical Theory and Applications* **16(3)**: 366-374.

Mahmoudi, M. R., Heydari, M. H., Avazzadeh, Z. (2018b). Testing the difference between spectral densities of two independent periodically correlated (cyclostationary) time series models, *Communications in Statistics -- Theory and Methods*, In Press.

Mahmoudi, M. R., Heydari, M. H., Avazzadeh, Z. (2018c). On the asymptotic distribution for the periodograms of almost periodically correlated (cyclostationary) processes, *Digital Signal Processing* **81:** 186-197.

Mahmoudi, M. R., Heydari, M. H., Roohi, R. (2018d). A new method to compare the spectral densities of two independent periodically correlated time series, *Mathematics and Computers in Simulation* **160**: 103-110.

Mahmoudi, M. R., Mahmoodi, M. (2014a). Inferrence on the Ratio of Variances of Two Independent Populations. *Journal of Mathematical Extension* **7(2)**:83-91.

Mahmoudi, M. R., Mahmoodi, M. (2014b). Inferrence on the Ratio of Correlations of Two Independent Populations. *Journal of Mathematical Extension* **7(4)**: 71-82.

Mahmoudi, M. R., Mahmoodi, M., Pak, A. (2019). On comparing, classifying and clustering several dependent regression models. *Journal of Statistical Computation and Simulation.* In Press.

Mahmoudi, M. R., Mahmoudi, M., Nahavandi, E. (2016). Testing the Difference between Two Independent Regression Models. *Communications in Statistics -- Theory and Methods* **45(21):** 6284-6289.

Mahmoudi, M. R., Maleki, M. (2017). A New Method to Detect Periodically Correlated Structure. *Computational Statistics* **32 (4)**: 1569-1581.

Mahmoudi, M. R., Maleki, M., Pak, A. (2017b). Testing the Difference between Two Independent Time Series Models. *Iranian Journal of Science and Technology: Sciences* **41(3):** 665-669.

Mahmoudi, M. R., Maleki, M., Pak, A. (2018e). Testing the Equality of Two Independent Regression Models, *Communications in Statistics-Theory and Methods* **47 (12)**: 2919-2926.

Maleki, M., Arellano-Valle, R. B., Dey, D. K., Mahmoudi, M. R., Jalali, S. M. J. (2017). A Bayesian approach to robust skewed autoregressive processes. *Calcutta Statistical Association Bulletin* **69 (2)**: 165-182.

Maleki, M., Contreras-Reyes, J. E., Mahmoudi, M. R. (2019). Robust Mixture Modeling Based on Two-Piece Scale Mixtures of Normal Family. *Axioms* **8 (2):** 38.

Maleki, M., Mahmoudi, M. R. (2017). Two-piece location-scale distributions based on scale mixtures of normal family. *Communications in Statistics-Theory and Methods* **46 (24)**: 12356-12369.

Naji, A., Kanaan, M., Ghassan, N. K., Bani, M., Basal, I. M. (2005). Statistical Classifier of the Holy QuranVerses (Fateha and Yaseen chapters), *Journal of Applied Science* **15(3)**: 580-583.

Nöldeke, T., (2012). *The History of the Qur'an*, Leiden: Brill, In Germany.

Sadeghi, B. (2011). The Chronology of the Qurān: A Stylometric Research Program, *Arabica* **58(3-4):** 210-299.

Suyuti, A. A. A. B. (1998), Alatghan Fi Olum El Quran, trans. by Haeri Ghazvini, Mehdi. Tehran. Amirkabir.

Yin, M. M., Mahmoudi, M. R., Abbasalizadeh, A. (2019). Analysis of mystical concepts in Khaghani's Divan. *Digital Scholarship in the Humanities*. In Press.

Samadianfard, Saeed, et al. "Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm." Energy Reports 6 (2020): 1147-1159.

Taherei Ghazvinei, Pezhman, et al. "Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network." Engineering Applications of Computational Fluid Mechanics 12.1 (2018): 738-749.

Qasem, Sultan Noman, et al. "Estimating daily dew point temperature using machine learning algorithms." Water 11.3 (2019): 582.

Mosavi, Amir, and Atieh Vaezipour. "Reactive search optimization; application to multiobjective optimization problems." Applied Mathematics 3.10A (2012): 1572-1582.

Shabani, Sevda, et al. "Modeling pan evaporation using Gaussian process regression K-nearest neighbors random forest and support vector machines; comparative analysis." Atmosphere 11.1 (2020): 66.

Ghalandari, Mohammad, et al. "Aeromechanical optimization of first row compressor test stand blades using a hybrid machine learning model of genetic algorithm, artificial neural networks and design of experiments." Engineering Applications of Computational Fluid Mechanics 13.1 (2019): 892-904.

Mosavi, Amir. "Multiple criteria decision-making preprocessing using data mining tools." arXiv preprint arXiv:1004.3258 (2010).

Karballaeezadeh, Nader, et al. "Prediction of remaining service life of pavement using an optimized support vector machine (case study of Semnan–Firuzkuh road)." Engineering Applications of Computational Fluid Mechanics 13.1 (2019): 188-198.

Asadi, Esmaeil, et al. "Groundwater quality assessment for sustainable drinking and irrigation." Sustainability 12.1 (2019): 177.

Mosavi, Amir, and Abdullah Bahmani. "Energy consumption prediction using machine learning; a review." (2019).

Dineva, Adrienn, et al. "Review of soft computing models in design and control of rotating electrical machines." Energies 12.6 (2019): 1049.

Mosavi, Amir, and Timon Rabczuk. "Learning and intelligent optimization for material design innovation." In International Conference on Learning and Intelligent Optimization, pp. 358-363. Springer, Cham, 2017.

Torabi, Mehrnoosh, et al. "A hybrid machine learning approach for daily prediction of solar radiation." International Conference on Global Research and Education. Springer, Cham, 2018.

Mosavi, Amirhosein, et al. "Comprehensive review of deep reinforcement learning methods and applications in economics." Mathematics 8.10 (2020): 1640.

Ahmadi, Mohammad Hossein, et al. "Evaluation of electrical efficiency of photovoltaic thermal solar collector." Engineering Applications of Computational Fluid Mechanics 14.1 (2020): 545-565.

Ghalandari, Mohammad, et al. "Flutter speed estimation using presented differential quadrature method formulation." Engineering Applications of Computational Fluid Mechanics 13.1 (2019): 804-810.

Ijadi Maghsoodi, Abteen, et al. "Renewable energy technology selection problem using integrated h-swara-multimoora approach." Sustainability 10.12 (2018): 4481.

Mohammadzadeh S, Danial, et al. "Prediction of compression index of fine-grained soils using a gene expression programming model." Infrastructures 4.2 (2019): 26.

Sadeghzadeh, Milad, et al. "Prediction of thermo-physical properties of TiO2-Al2O3/water nanoparticles by using artificial neural network." Nanomaterials 10.4 (2020): 697.

Choubin, Bahram, et al. "Earth fissure hazard prediction using machine learning models." Environmental research 179 (2019): 108770.

Emadi, Mostafa, et al. "Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran." Remote Sensing 12.14 (2020): 2234.

Shamshirband, Shahaboddin, et al. "Developing an ANFIS-PSO model to predict mercury emissions in combustion flue gases." Mathematics 7.10 (2019): 965.

Salcedo-Sanz, Sancho, et al. "Machine learning information fusion in Earth observation: A comprehensive review of methods, applications and data sources." Information Fusion 63 (2020): 256-272.