Journal of Engg. Research Vol.9 No. (4B) December 2021 pp. 211-233 DOI:10.36909/jer.11035

Spatial and temporal multivariate statistical analysis to assess drinking water quality in medical services

Abdalrahman Alsulaili* and Sarah Alshawish**

** Department of Civil Engineering, Kuwait University, PO Box 5969 Safat, Kuwait*
*** Department of Chemical Engineering, Kuwait University, PO Box 5969 Safat, Kuwait*
** Corresponding Author: a.alsulaili@ku.edu.kw*

Submitted ***:*** 04/07/2020
Revised ***:*** 18/10/2020
Accepted ***:*** 28/10/2020

ABSTRACT

Drinking water quality supplied to medical services presents significant role regarding the health aspect of the society. Multivariate statistical techniques were applied for the interpretation of data obtained, i.e., cluster analysis (CA), principal component analysis (PCA), factor analysis (FA), and discriminant analysis (DA) to analyze and assess the spatial and temporal variations of drinking water quality in different medical services in Kuwait. This study was generated over a period of 11 years (2007–2017), including 19 parameters at fourteen different sites. Hierarchical CA obtained two groups regarding both spatial and temporal variations. For spatial variations, 14 sampling sites were grouped into Low Concentration (LC) and High Concentration (HC). For temporal variations, 12 months were grouped into Summer and Winter. DA provided better results by data reduction for the large data set with great discriminatory ability for both spatial and temporal variations, as only five parameters were used concerning the spatial variations to afford 68.4% of the cases being assigned correctly, and seven parameters were interpreted for the temporal variations affording 76.1% of correctly classified cases. The applied PCA/FA on the spatial variations resulted in five principle components (PCs) for the LC region, and the total variance is 74.84% and three PCs for the HC region explaining a total variance of 64.86%. For the temporal variations, summer yielded into five PCs with a total variance of 70.6%, whereas the winter resulted in three PCs describing 67.1% total variance. Thus, multivariate analysis provides better spatial and temporal variations assessment in contemplation of effective drinking water quality management and control.

Keywords: Cluster analysis; Discriminant analysis; Drinking water quality; Factor analysis; Medical services; Principle component analysis.

1. INTRODUCTION

Drinking water is a primary proposition that needs to be observed, by the reason of the unprecedented growth of the population that may increase pollution along with other factors such as industrialization and urbanization (Velea et al., 2009). Water resources contamination has a significant impact on the environment and human health (Emmanuel et al., 2009; Muhammad et al., 2011). For example, the anthropogenic sources caused a worldwide concern, by which drinking water is contaminated with various chemicals and heavy metals (Rapant & Krčmová, 2007). Thus, water quality and cleanliness accessible in hospitals are considered to be one of the most important issues to concentrate on. Because of that, awareness must be spread and directed towards prioritizing the safety of water, since hospitals compared to different types of institutions, utilizes high loads of water on a daily basis for different kinds of applications such as ingestion of water, personal hygiene, washing surgical tools and kidney dialysis. Ojo & Adeniyi (2012) acknowledged that the average usage of water in a regular hospital is 100 liters/person/day, whereas the water supplied to hospitals ranges between 400 to 1200 liters/day/bed. Thus, monitoring the quality of water and assuring the execution of guidelines set will lead to improving the performance of the hospitals and other medical facilities along with upgrading its health service.

Several studies have been made related to the assessment of water quality using multivariate statistical techniques regarding river basins, as a study was applied concerning the evaluation of water quality data of both spatial and temporal variations in Gomti River (India) (Singh et al., 2005). Adding to that, Barakat et al. (2016) conducted a study where the main focus was on determining the contamination sources in the river using multivariate statistical techniques by assessing the spatial and seasonal water quality of Oum Er Rbia River in Morocco. Another study was applied on Muda River basin, to analyze its water quality based on a data matrix generated through several years, having various multivariate techniques being subjected to them (Azhar et al., 2015). Regarding the studies applied on lakes, Kazi et al. (2009) operated several multivariate statistical techniques in order to analyze pollution sources and design a monitoring network. Similarly, Y. Zhao et al. (2012) studied the sources of pollution along with interpreting and evaluating the data of water quality, running the same multivariate techniques for these operations. Previous studies have shown good results in characterizing and evaluating the data using multivariate analysis. Thus, these techniques have been proven to be a valid method in classifying water properties. Moreover, a thorough literature review was conducted with limited studies discussing the same matter regarding the analysis of drinking water in different medical services using the multivariate techniques were found (Yu et al., 2008). The applied multivariate statistical techniques in this study which are cluster analysis (CA), principle component/factor analysis (PCA/FA) and discriminant analysis (DA), are essentially used in order to minimize the number of variables having the relationships between the original data retained, which aims into identifying the most significant factors that are responsible for most of the discrepancies existing in the process (Ouyang et al., 2006; Shrestha & Kazama, 2007).

The objective of this study is to analyze the large data of drinking water quality parameters cumulated in a duration of 11 years (2007–2017) on several medical services around Kuwait, highlighting several physical, chemical and biological parameters such as temperature, pH, total dissolved solids, total hardness, chloride, calcium and total coliform. That is done by applying different multivariate statistical techniques, the results are expected in classifying the parameters that are accountable for the quality of drinking water variations among the different monitoring sites as well as the acknowledgement of the differences existing between these parameters. Notably, the water samples were taken from distribution lines directly before the medical centers.

2. MATERIALS AND METHODS

2.1 Study Area

Kuwait has a total area of 17,818 square Kilometers (Kuwait government, 2018); however, 88% of the total area is observed as desert, the rest 12% is considered to be as an urban area, where the analysis took place. The data used in this study covers the period between January 2007 until December 2017 and was obtained in collaboration with the Ministry of Electricity and Water (MEW) and Kuwait Environmental Protection Agency (KWEPA). Field samples were collected, on a monthly basis, from 14 different sites around the urban area covering its northern and southern parts Figure 1, for the utilization of statistical analysis on the parameters to be investigated. These samples are collected to obtain and study the yearly statistics of water quality supplied to medical services around Kuwait in order to maintain the characteristics of water within the accepted limitations by the World Health Organization (WHO) and KWEPA and report for any possible issues that could occur. Moreover, this study targeted not only hospitals but covered as well other aspects of medical services such as health centers and clinics around the six districts of the country. Al Ahmadi district includes Al Fahaheel Clinic (site 1), Al Qurain Clinic (site 4), Al Riqqa Health Centre (site 5), Ali Sabah Al Salem Clinic (site 8). As for Farwaniya district, it includes Al Farwaniya Hospital (site 2) and Al Omariya Health Centre (site 12). Regarding Hawalli district, it involves four sites which are, Al Salmiya Clinic (site 6), Bayan Health Centre (site 9), Western Hawalli Health Centre (site 10), Rumaithiya Medical Clinic (site 13). For the Capital district, two sites are included, Al Sulaibikhat Clinic (site 7) and Sharq Hospital (site 14). Each of Al Jahra and Mubarak Al Kabeer Districts includes one site only, Al Jahra Hospital (site 3) and Mubarak Al Kabeer Health Centre (site 11), respectively. All the sites investigated belongs to the government sector.

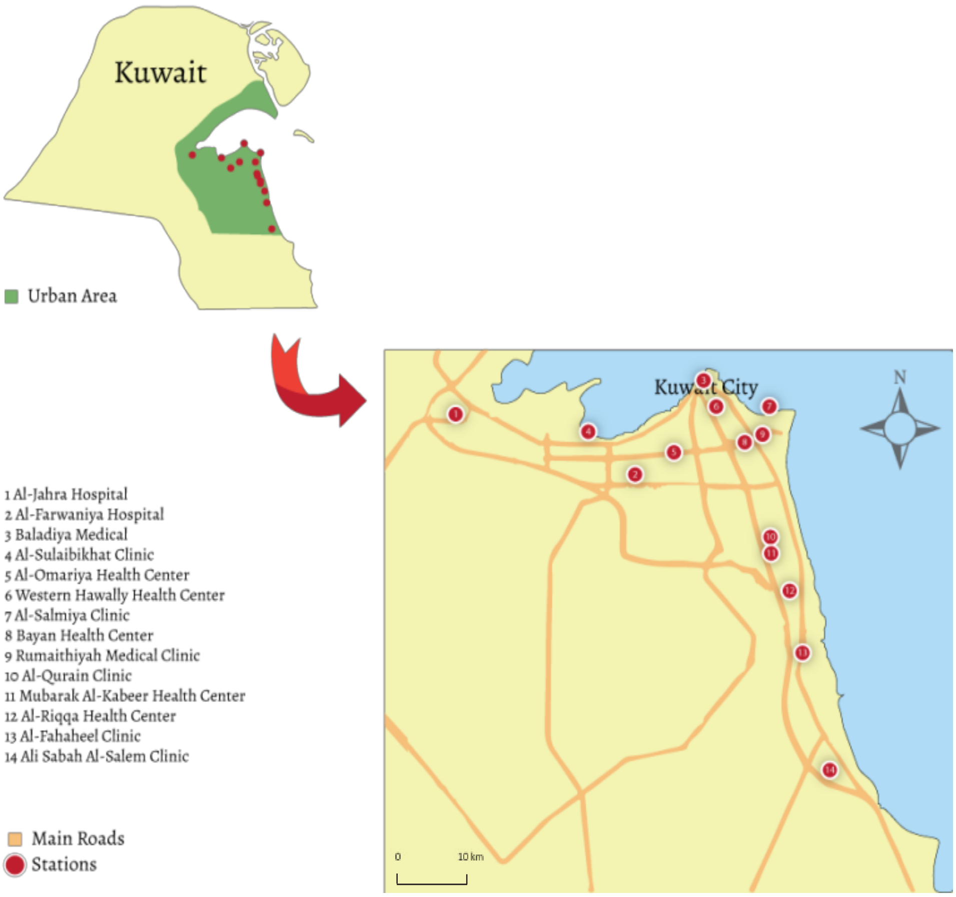


Figure 1: The map of Kuwait and the drinking water quality stations.

2.2 Drinking Water Quality Data and Analytical Procedures

The data set amount to a total of forty-three drinking water quality parameters. However, due to the large amount of missing data (above 50% missing) related to certain parameters (24 parameter), the analysis in this study was made on nineteen parameters listed in Table 1 along with their units and analytical methods. Moreover, basic descriptive analysis was done on each variable for the entire period of study and summarized in Table 2.

Table 1: Water quality parameters, units and analytical methods used between 2007 - 2017 for Kuwait medical services.

Parameters	Abbreviations	Units	Analytical Methods
Temperature	T	°C	Mercury Thermometer
Electrical Conductivity	EC	µS/cm	Ionic Chromatography
Total Dissolved Solids	TDS	mg/l	Ionic Chromatography
pH	pH	0 – 14	pH-meter
Chloride	Cl	mg/l	Ionic Chromatography
Sodium	Na	mg/l	Ionic Chromatography
Sulphate	SO4	mg/l	Ionic Chromatography
Calcium	Ca	mg/l	Ionic Chromatography
Magnesium	Mg	mg/l	Ionic Chromatography
Potassium	K	mg/l	Ionic Chromatography
Total Hardness	TH	mg/l	Ionic Chromatography
Total Alkalinity	TA	mg/l	Ionic Chromatography
Bromide	Br	mg/l	Ionic Chromatography
Fluoride	F	mg/l	Ionic Chromatography
Nitrate	NO3	mg/l	Ionic Chromatography
Free Residual Chlorine	FRC	mg/l	Ionic Chromatography
Fecal Coliform	FC	CFU/100ml	Membrane Filtration System
Total Coliform	TC	CFU/100ml	Membrane Filtration System
Pseudomonas Aeruginosa	PA	CFU/250ml	Membrane Filtration System

Table 2: Range, mean, and S.D. of different water-quality parameters at different sites of medical services in Kuwait between 2007 and 2017.

Parameters	WHO Limits	EPA-KW	EPA-US	Descriptive Analysis	Al Fahaheel Clinic	Al Farwaniya Hospital	Al Jahra Hospital	Al Qurain Clinic	Al Riqqa Health Center	Al Salmiya Clinic	Al Sulaibikhat Clinic	Ali Sabah Al-Salem Clinic	Bayan Health Center	Hawalli Health Center	Mubarak Al-Kabeer Health Center	Omariya Health Center	Rumaithiya Medical Clinic	Sharq Hospital
Temperature ( )	-	-	-	Range	15 - 40	18 - 41	16 - 48	18 - 37	17 - 46	19 - 38	13 - 49	14 - 39	10 - 40	18 - 43	21 - 28	16 - 31	20 - 34	18 - 40
				Mean	26.408163	27.158537	30.04902	23.517647	28.367347	29.731707	28.663043	26.715686	26.194175	26.754902	24.8	23.842105	25.263158	27.148515
				Std. Dev.	4.58783	5.834062	7.871967	3.517344	6.820811	4.799817	7.353368	4.527971	5.942631	5.567651	2.529822	3.609604	4.214289	4.821589
EC (µS/cm)	-	-	-	Range	86.7 - 616	124.7 - 637	110.5 - 721	3.8 - 464	70 - 588	105 - 561	1.05 - 707	87 - 474	69.3 - 555	138 - 673	21.2 - 382.3	182.3 - 432	184.2 - 326.8	151.1 - 675
				Mean	255.546226	329.276087	325.046082	235.904255	256.565714	297.41087	347.807386	232.770909	290.805405	322.70367	256.088889	281.347059	249.082353	357.385321
				Std. Dev.	105.215544	132.823241	157.06586	91.879299	91.5195	121.734755	156.670123	80.260595	101.871214	143.423086	103.031554	73.197918	33.099041	146.85881
TDS (mg/l)	1000	1000	500	Range	55 - 394.2	79.8 - 421	70.7 - 461.44	46.2 - 328	44.8 - 376.32	67.2 - 376	111.8 - 452.48	55.7 - 303.3	44.3 - 365	88.3 - 477	13.57 - 244.48	116.67 - 276.48	117.88 - 209.15	96.7 - 510
				Mean	170.206411	219.515217	215.02784	159.568842	168.238208	198.948261	233.235596	153.756161	192.906464	213.499436	156.775333	178.967444	160.517444	235.241838
				Std. Dev.	70.541512	90.300776	105.653815	60.424582	59.420756	82.238044	101.281983	53.365894	69.843989	97.411954	61.645782	45.683546	21.076067	101.190454
pH	6.5 - 8.5	6.5 - 8.5	6.5 - 8.5	Range	6.2 - 8.62	6.47 - 8.44	6.25 - 8.5	6.9 - 8.8	6.68 - 8.69	6.6 - 8.27	6.4 - 8.07	6.7 - 8.65	6.68 - 8.7	6.5 - 8.3	7.05 - 8.47	6.8 - 8.3	6.95 - 8.23	6.71 - 8.7
				Mean	7.652303	7.675451	7.585359	7.882097	7.783653	7.662198	7.534831	7.755788	7.776913	7.651214	7.734444	7.411818	7.455455	7.697087
				Std. Dev.	0.405734	0.353958	0.401159	0.41142	0.421795	0.335711	0.358639	0.393846	0.362967	0.370335	0.457469	0.402165	0.380956	0.382317
Chloride (mg/l)	250	250	250	Range	6.54 - 109.08	7.93 - 211.3	7.91 - 119.9	6.07 - 119.9	6.69 - 159.9	9.59 - 147.59	9.99 - 139.9	4.27 - 137	6.02 - 119.9	9.99 - 129.9	12.06 - 131	18 - 101.5	24.4 - 126	11.9 - 179.9
				Mean	48.692944	53.068604	43.977993	44.746313	49.287974	49.200041	51.530801	44.299982	49.646991	49.194285	51.368	49.973778	50.276889	59.726682
				Std. Dev.	29.203728	34.946976	28.467683	27.616467	29.227975	30.267126	30.934292	29.390581	28.802206	29.672859	35.649008	25.824174	25.72491	32.733145
Sodium (mg/l)	200	200	-	Range	4.29 - 59.15	4.06 - 156.4	4.1 - 87.99	4.36 - 63.34	5.09 - 74.6	5.83 - 66.41	7.84 - 84.99	2.8 - 58.04	5.56 - 60.74	5.82 - 76.8	7.44 - 44	10 - 50	12.8 - 52.8	7.68 - 84.61
				Mean	24.078993	29.512208	26.504728	23.833477	24.392529	26.686703	28.962651	22.618139	25.440286	27.426879	21.62	22.075111	21.710944	32.950137
				Std. Dev.	13.067751	20.08163	17.722236	13.342965	12.5485	14.726649	17.27918	12.545453	12.717025	14.904454	11.012582	10.020407	9.56712	14.89042
Sulphate (mg/l)	250	250	250	Range	0.01 - 105.82	5.49 - 144.46	0.58 - 184.9	1.52 - 107.2	0.19 - 113.3	2.51 - 158.81	2.79 - 187.4	0.01 - 136.2	0.93 - 114.67	2.17 - 178.47	0.02 - 6.07	1.26 - 51	0.5 - 16	0.9 - 184.86
				Mean	24.005841	48.655455	48.452465	24.15223	22.381929	44.520327	56.414689	18.470736	29.421645	46.371427	1.9252	8.583944	8.207278	47.316043
				Std. Dev.	26.202815	38.656035	49.41195	24.898102	24.546828	38.31796	50.691553	23.18699	26.909546	50.962574	1.904243	11.37098	4.485077	54.77785
Calcium (mg/l)	-	200	-	Range	6.97 - 56.11	12.38 - 60.7	4.77 - 68.85	7.41 - 54.5	8.22 - 64.12	10.92 - 64.12	9.58 - 68.94	7.67 - 40.64	5.34 - 52.9	6.96 - 64.06	12.04 - 38.7	14.25 - 27.65	16.2 - 24.23	10.9 - 69.81
				Mean	20.071778	29.914597	29.744881	20.835462	22.098588	25.946191	33.306345	20.082831	26.835399	29.493667	24.9867	20.915944	20.227222	29.734415
				Std. Dev.	9.706834	12.14356	13.181553	8.365583	9.42596	11.122137	13.184899	5.850372	10.071712	12.714662	7.481803	3.500429	2.116422	13.339024
Magnesium (mg/l)	-	150	-	Range	0.05 - 9.94	0.15 - 24.51	0.27 - 16.99	0.02 - 7.72	0.1 - 7.94	0.12 - 12.9	0.21 - 17.06	0.03 - 9.19	0.02 - 10.23	0.1 - 15.66	0.11 - 4.7	0.36 - 5.23	1.48 - 7.6	0.04 - 17.97
				Mean	2.8385	4.696901	4.554837	2.34303	2.558204	3.796051	5.201394	2.23522	3.597426	4.125964	2.3587	2.617222	3.248889	4.219748
				Std. Dev.	2.164201	4.216725	4.009133	1.620546	1.507456	3.08151	4.288705	1.573432	2.516941	3.81963	1.320684	1.411075	1.73718	4.210167
Potassium (mg/l)	-	10	-	Range	0.01 - 5.45	0.01 - 4.89	0.04 - 2.96	0.1 - 4.44	0.02 - 15.02	0.02 - 3	0.01 - 2.73	0.01 - 4.4	0.01 - 4.65	0.2 - 3.1	0.43 - 3.9	0.54 - 5	0.87 - 4.3	0.26 - 8.8
				Mean	1.470454	1.190898	1.211329	1.339267	1.798126	1.14734	1.30122	1.527567	1.306605	1.365825	2.0304	2.077167	2.282278	1.611517
				Std. Dev.	1.1007	0.773392	0.723039	0.881559	1.784188	0.641532	0.62734	1.060311	0.837111	0.66686	1.135865	1.024737	0.796779	1.032705
Total Hardness (mg/l)	-	500	-	Range	0 - 280	0 - 280	0 - 241.1	0 - 180	0 - 200	0 - 280	0 - 280	0 - 172	0 - 369.9	0 - 236	30.53 - 110.41	37.06 - 80.3	52 - 82.7	0 - 260
				Mean	70.988354	109.575954	104.823897	69.520413	72.900486	94.524084	122.207132	68.183923	91.559085	101.469153	72.1022	64.230467	65.715467	104.830874
				Std. Dev.	46.963737	64.938244	60.64694	36.243379	38.234019	54.96649	71.52477	32.347283	50.859223	59.663509	21.486938	13.210267	10.229726	64.431842
Total Alkalinity (mg/l)	-	100	-	Range	12 - 56	16 - 140	12 - 80	20 - 80	16 - 120	16 - 80	16 - 68	19.2 - 80	16 - 92	16 - 80	8 - 76	12 - 64	16 - 48	20 - 100
				Mean	32.43301	42.248864	40.157944	37.526374	39.707843	39.388636	42.425843	38.881481	42.042593	44.360377	48.444444	39.944444	37.166667	47.590654
				Std. Dev.	8.90398	15.474621	11.628016	11.066866	16.443764	11.038047	10.229675	14.025852	13.703456	12.282498	20.143927	12.63735	7.302296	12.406066
Bromide (mg/l)	-	-	-	Range	0 - 1.02	0.01 - 2.51	0 - 3.51	0 - 0.69	0 - 1.06	0.01 - 0.77	0.01 - 0.94	0 - 1.38	0 - 1.63	0 - 0.89	0.01 - 0.97	0.01 - 0.65	0.01 - 0.84	0.01 - 1.4
				Mean	0.119162	0.147074	0.173208	0.114678	0.131441	0.113844	0.113577	0.148713	0.116548	0.112335	0.2788	0.090056	0.097529	0.152286
				Std. Dev.	0.174146	0.332465	0.502428	0.156116	0.210093	0.148818	0.178569	0.240127	0.214075	0.165922	0.396406	0.150636	0.196806	0.211411
Fluoride (mg/l)	1.5	1.5	2 - 4	Range	0 - 0.72	0.01 - 0.54	0 - 1.59	0 - 0.44	0 - 0.79	0 - 0.53	0.01 - 0.58	0 - 0.2	0.01 - 1.1	0 - 0.83	0 - 0.16	0.01 - 0.15	0 - 0.08	0 - 0.57
				Mean	0.072997	0.106618	0.122387	0.072188	0.076386	0.092681	0.109936	0.065581	0.082095	0.098848	0.029143	0.024923	0.018833	0.113559
				Std. Dev.	0.091755	0.092574	0.17775	0.075389	0.109192	0.087039	0.116439	0.053139	0.121921	0.109842	0.058044	0.037951	0.020275	0.117379
Nitrate (mg/l)	50	50	10	Range	0 - 3.61	0.01 - 4.55	0 - 4.79	0 - 2.88	0 - 12.13	0.01 - 4.08	0.01 - 4.66	0 - 3.89	0 - 5.05	0.01 - 4.84	0 - 1.01	0 - 1.08	0.01 - 0.96	0 - 5.54
				Mean	0.675442	1.226501	1.079118	0.569471	0.657537	1.038541	1.139575	0.46561	0.9635	1.001368	0.302333	0.340722	0.372118	0.930377
				Std. Dev.	0.820397	1.046512	1.063967	0.572079	1.355087	0.90846	1.032772	0.628065	0.971866	1.012166	0.35155	0.292232	0.297216	1.006585
Free Residual Chlorine (mg/l)	0	0	0	Range	0 - 0.1	0 - 0.4	0 - 0.5	0 - 0.2	0 - 0.2	0 - 0.2	0 - 0.3	0 - 0.2	0 - 0.3	0 - 0.4	0	0 - 0.2	0	0 - 0.3
				Mean	0.01	0.105319	0.044737	0.012371	0.023853	0.026596	0.048913	0.023684	0.023478	0.051754	0	0.021053	0	0.053097
				Std. Dev.	0.030137	0.114879	0.082109	0.038885	0.052527	0.049024	0.074855	0.050313	0.061173	0.081178	0	0.06306	0	0.075676
Faecal Coliform (CFU/100ml)	0	0	0	Range	0	0	0	0	0	0	0	0	0	0	0	0	0	0
				Mean	0	0	0	0	0	0	0	0	0	0	0	0	0	0
				Std. Dev.	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TotalColiform (CFU/100ml)	0	0	0	Range	0 - 30	0	0	0	0	0	0	0	0 - 1	0	0	0	0	0 - 1
				Mean	0.272727	0	0	0	0	0	0	0	0.008696	0	0	0	0	0.008929
				Std. Dev.	2.860388	0	0	0	0	0	0	0	0.09325	0	0	0	0	0.094491
PseudomonasAeruginosa (CFU/250ml)	0	0	0	Range	0 - 141	0	0	0	0	0	0	0	0	0	0	0	0	0
				Mean	1.453608	0	0	0	0	0	0	0	0	0	0	0	0	0
				Std. Dev.	14.316381	0	0	0	0	0	0	0	0	0	0	0	0	0

2.3 Data Treatment

The nonparametric correlation, Spearman rank's correlation coefficient is used as a measurement of strength association between the ranked variables as well as its direction. In this study, the correlation had been accountable for the water quality parameters of which being non-normally distributed (Alberto et al., 2001).

Kaisere-Meyere-Olkin (KMO) and Bartlett's test were performed in order to analyze he suitability of the data regarding the principle component analysis and factor analysis. The purpose of KMO test is the measurement of sampling adequacy which shows the percentage of variance due to the underlying factor in the variables used in this study. Values that are close to 1 generally indicates that the sampling is adequate, resulting in positive and useful impact on the data. However, values that are below 0.5 are impractical with the data (Kaiser, 1974; Merkle et al., 1998; Sabatini, 2002). In this study, concerning the spatial variations, for LC and HC the KMO test resulted in the values of 0.756 and 0.841 respectively. As for the temporal variations for summer and winter, the values were 0.843 and 0.823 respectively. Regarding Bartlett's test of sphericity, it was applied to examine the null hypothesis that shows the correlation matrix to be an identity matrix, resulting into uncorrelated variables. A level of significance less than 0.05 indicates suitable fit of data (Hair et al. 1995; Tabachnick & Fidell, 2001). In our study, the null hypothesis was rejected, as the significance value is 0. This determines an efficient application of the PCA/FA on the data set.

3. MULTIVARIATE STATISTICAL ANALYSIS

Multivariate statistical techniques are used on a wide range of different fields such as quality control, research and development, process optimization and market research. In this study, it was applied in order to classify and describe the properties of water quality along with analyzing both spatial and temporal variations. The multivariate analysis utilized in this study was performed using IBM SPSS Statistics 26. The drinking water quality data sets were interpreted by running three different multivariate statistical techniques.

3.1 Cluster Analysis (CA)

The main objective of this approach is to classify the cases into a number of clusters (groups) based on their characteristics and similarities within the group and dissimilarities among different groups. In this study, hierarchical agglomerative clustering was applied at which homogeneous clusters are formed by combining cases in such a way where they merged together one by one in a continuously ordered steps, and it is generally presented by a dendrogram (tree diagram) (Blei & Lafferty, 2009). CA significance is determined by the $0.66D_{max}$ criterion (Lovchinov & Tsakovski, 2006). Furthermore, hierarchical CA was applied using Ward's method with squared Euclidean distances as a measure of similarity (Singh et al., 2005; J. Zhao et al., 2011) between two samples and a distance can be represented by the difference between analytical values from the samples.

CA was applied to the drinking water quality data set based on spatial and temporal variation. For the spatial variation, CA is performed to identify similar site locations (medical services) distributed across Kuwait. For the temporal variation, CA is performed on the data set to identify similarity based the months of the year. The result is shown as a dendrogram and linkage distance is reported as D_{link} / D_{max} (SPSS scales the linkage distance to 0 – 25) (Singh et al., 2004).

3.2 Discriminant analysis

Discriminant Analysis classifies cases of data into categories based on the contrast in the values of their attributes. The difference between DA and CA lies in that DA requires predetermined categories. If DA is effective for a data set, the results will show a high correct percentage for each assigned group. In DA, a discriminant function for each group is built as in the following equation (Singh, Malik, & Sinha, 2005):

$$f(G_k) = C_k + \sum_{i=1}^{n} w_{ik} p_{ik} \quad (1)$$

where k is the number of predetermined groups (G), C_k is the constant for each group, n is the number of parameters/attributes (p_i) used to contrast between the given groups, w_{ik} is the weight assigned to each parameter in each group. The weight coefficient is estimated to maximize the difference between the means of the selected parameters. The DA results is shown as a classification table, which is used to assess the performance of the DA and the validity of the assigned groups. The classification table has the observed categories as rows and predicted categories as columns. Thus, the correct classification cases are shown on the diagonal.

In this study, two spatial and two temporal groups were selected for the DA. The two spatial groups were selected based on the cluster analysis results, and 17 parameters were selected as the criteria for classifying these groups. Standard DA, forward stepwise DA, and backward stepwise DA were performed on both the spatial and temporal variation. The spatial variation consisted of the hospitals while the temporal variation consisted of seasons, and both are the dependent variable in the DA. The independent variables included the measured water quality parameters.

3.3 Principle Component Analysis (PCA)

Principal component analysis is used to point out the significant parameters, which describes the data set. Thus, PCA allows for data reduction without significant loss of data behavior and characteristic (Helena et al., 2000). The principal component can be given by

$$z_{ij} = c_{i1}x_{1j} + c_{i2}x_{2j} + \cdots + c_{in}x_{nj} \quad (2)$$

where z is the component score, i is the component number, j is the sample number, c is the loading of the component, x is the measured variable value, and n is the total number of variables. Furthermore, PCA provides linear combination, called principal components, of the original variables. In this study, PCA were used to extract significant principal components and reduce the data set while maintaining its structure.

4. DISCUSSION AND RESULTS

4.1 Spatial Similarity (hospital grouping)

To detect the similarity between medical services (spatial similarity) cluster analysis (CA) was applied on all the 14 medical services, except Mubarak Al-Kabeer Health Center. This health center was excluded as it has no values measured for PS. Figure 2 shows the dendrogram generated by the spatial cluster analysis. Medical services were grouped into two statistically significant clusters at $(D_{link}/D_{max}) \times 100 < 40$ (SPSS rescale the linkage distance to fall in the range of 1 to 25). Dlink is the linkage distance for a particular case and the Dmax is the maximum linkage distance. In Hierarchical cluster analysis, the number of clusters is decided based on practical information. Thus, the number of clusters was decided based on the mean values of drinking water characteristics into Low Concentration (LC) and High Concentration (HC). Cluster 1, the low concentration (LC) cluster, includes Al Fahaheel Clinic (site 1), Al Qurain Clinic (site 4), Al Riqqa Health Center (site 5), Ali Sabah Al Salem Clinic (site 8), Omariya Health Center (site 12), and Rumaithiya Medical Clinic (site 13). Geographically, LP hospitals are located in the southern region of Kuwait residential areas, except for Omariya Health Center (site 12) and Rumaithiya Medical Clinic (site 13).

As for Cluster 2, HC medical services, includes Al Farwaniya Hospital (site 2), Al Jahra Hospital (site 3), Al Salmiya Clinic (site 6), Al Sulaibikhat Clinic (site 7), Bayan Health Centert (site 9), Hawalli Health Center (site 10), and Sharq Hospital (site 14). HC medical services are all located in the northern region of Kuwait residential areas. Cluster analysis results suggest the water sources and distribution system in the south provides medical services with water of low concentration values than water sources and distribution systems in the north. It should be noted that the age of the medical services has been studied; however, no significant classification could be interpreted from it. However, (Liu et al., 2011) conducted a study concerning the water quality in Yuan-Yang Lake that is a Subtropical Alpine Lake located in north-central Taiwan. The study area consisted of 8 sampling sites, that resulted into three significant clusters, as cluster 1, 2 and 3 represented the middle water depth, shallow area and deep zone, respectively. Moreover, the case study based on (X. Wang et al., 2012) targeted 12 sampling sites of the Xiangxi river basin in China. CA generated three statistically meaningful clusters which are low pollution upper catchments region (US), high pollution lower catchments region (LS) and moderate pollution middle catchments region (MS).

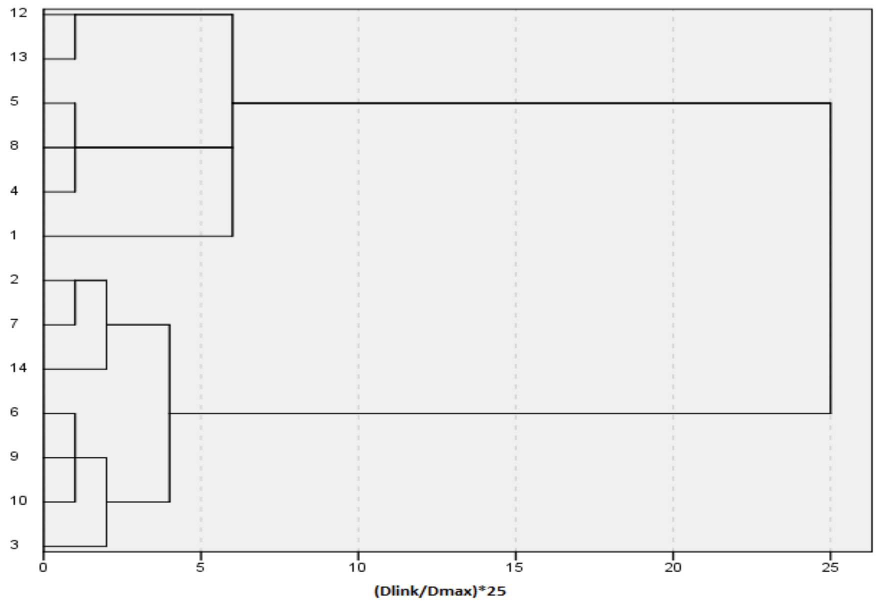


Figure 2: Dendrogram showing clustering of medical services based on drinking water characteristics.

4.2 Temporal Similarity (months grouping)

To study the temporal variability, cluster analysis was applied to drinking water quality parameters which were divided by the months of the year (12 months). The resulting dendrogram, shown in Figure 3, indicates that there are two statistically significant clusters at $(D_{link}/D_{max}) \times 100 < 80$. Cluster 1 includes months from June to October, in addition to April. While Cluster 2, includes months from November to February, in addition to May. Temporal Cluster analysis indicates that drinking water quality parameter can be grouped into two main seasons, Summer (cluster 1) and Winter (Cluster 2). This result was expected, as Kuwait is known for two main seasons (Summer and Winter) with April and May being the transition months. Similarly, (Muangthong & Shrestha, 2015) applied a case study on Nampong River and Songkhram River in Thailand, as the clustering procedure of both rivers generated two temporal groups, which represents two seasons, wet and dry.

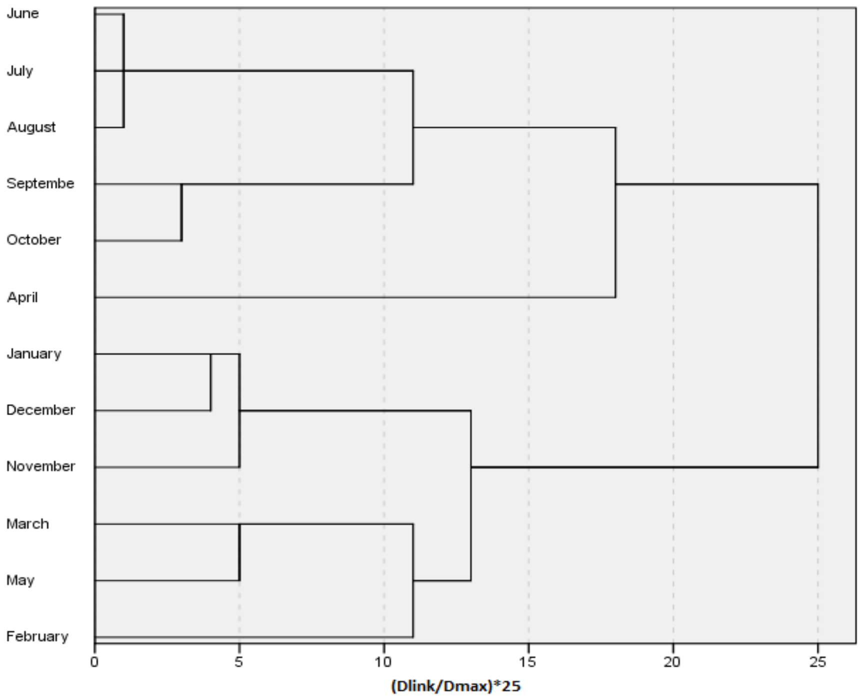


Figure 3: Dendrogram showing clustering of months based on drinking water characteristics.

4.3 Discriminant Analysis

DA was done based on the groups yielded in the CA (Spatial and Temporal). This analysis was conducted including only seventeen out of nineteen parameters, due to the lack of data regarding the two parameters [Faecal Coliform – Pseudomonas Aeruginosa]. From this test, the main focus was to attain the Discriminant functions (DFs) and Classification Matrices (CMs) from three different modes of the DA which are, standard, forward stepwise and backward stepwise modes as shown in Tables 3 and 4.

Each mode follows a particular order, in which parameters are added into the analysis, and, as for the forward stepwise mode, the most significant variables are added step-by-step until no significant changes are attained, whereas, in the backward stepwise mode, the least significant variables are eliminated step-by-step until no significant changes are attained. However, for the standard mode, all variables are added at once with the least significant being eliminated.

Regarding the spatial variations, the DA applied in the standard mode used 16 discriminant variables for the DFs that resulted in the corresponding CMs with 74.4% of original grouped cases were correctly classified. For the backward stepwise mode, eight discriminant variables were included for the DFs giving CMs with 76.96 % of cases being correctly classified. However, for the forward stepwise mode, the analysis gave corresponding CMs with 68.41% of correct cases assigned using only five discriminant variables Table 5. Forward stepwise analysis showed that Na, Ca, Total Hardness, Total Alkalinity, and Free Residual Chlorine, are the most significant parameters. Cl, Bromide, and pH are less significant than the previous group of parameters. Moreover, the last group of the 9 remaining parameters are much less significant as shown from standard mode DA. Thus, DA indicates that Na, Ca, Total Hardness, Total Alkalinity, and Free Residual Chlorine are responsible for spatial variations in drinking water quality. (Shrestha & Kazama, 2007) conducted the same analysis, using 12 parameters for the standard and forward stepwise method resulting in 83% of the cases being correctly classified, however for the backward stepwise method, only 7 parameters were included with 81% of the cases being assigned correctly. Similarly, Singh et al. (2004) applied DA on 24 parameters, nonetheless only 23 of them were used in the standard stepwise method while 17 of the parameters were included in the forward stepwise method, however both of the methods had 92% of the cases being correctly classified. As for the backward stepwise method, it consisted of 5 parameters only being the most significant and having 91% of the cases being assigned correctly.

For the parameters determined by the spatial DA using backward stepwise method, box, and whisker plots Figure 3 were constructed to evaluate different patterns associated with spatial variations in drinking water quality of several medical services. For spatial variations, higher concentrations of calcium, sodium, Total Hardness, and Total Alkalinity were found in the HC hospitals higher than LC hospitals. This can be related to water intake and the performance of the desalination plants which vary according to location.

As for the temporal variations, 16 discriminant variables were used in the standard mode giving CMs with 79.12% of correctly classified cases. For both forward and backward stepwise mode, only eight discriminant variables were used as shown in Table 6 resulting in corresponding CMs of 76.14% and 76.59% of the cases being correctly assigned, respectively. Both forward and backward stepwise methods indicates that there eight parameters [Temperature, pH, K, Total Alkalinity, Bromide, Fluoride, Nitrate, and Free Residual Chlorine] which are discriminating parameters in seasonal variation. Temporal variations were conducted in previous studies, in which Singh et al. (2005) applied the DA on 31 parameters concerning the standard stepwise analysis and 22 parameters with respect to the forward stepwise method, as both resulted in 97% of correctly assigned cases. Only 5 parameters regarding the backward stepwise method were accounted in the analysis and the cases that were correctly classified gave a value of 94%. Moreover, in Zhou et al. (2007) 23 parameters were involved in the temporal variation DA, as for both standard and forward stepwise methods, 23 and 13 parameters resulting in the corresponding CMs with 92% and 91.45% of the original grouped cases being correctly classified. Furthermore, 7 discriminant parameters only were included in the backward stepwise method with a 90.65% of correctly assigned cases.

Different seasonality patterns in parameters were evaluated by constructing box and whisker plots Figure 4 of the significant parameters given by the DA. The average temperature in Summer were found to be higher in summer and lower in Winter, which is expected. Also, a pH average value was found to be slightly higher in Summer than winter, however, the average pH values is under the standard limit for drinking water. Potassium and fluoride were higher in winter and lower in summer. Total Alkalinity in summer is higher than that in winter. The inverse relationship between total alkalinity and potassium concentrations is explained by the use of potassium chloride as water softener. Furthermore, when total alkalinity is low, excess potassium chloride is used to treat water, and thus increasing the potassium concentration. Nitrate Concentration was found higher in summer than winter, which can be due to the increase in the usage of agriculture fertilizers in arid climate in the summer.

For both, spatial and temporal variation, Free Residual chloride was found in winter and in HC sites only. This concentration is due to the disinfection process which leads to excess chloride in winter due to low microorganism rates.

Table 3: Classification matrix for discriminant analysis of spatial variations in the medical services

Monitoring Regions	% Correct	Regions Assigned by DA	
		LC	HC
Standard DA mode			
LC	69.9	186	80
HC	86.6	44	284
Total	79.1	266	328
Forward stepwise DA mode			
LC	55.9	227	179
HC	76.2	154	494
Total	68.4	406	648
Backward stepwise DA mode			
LC	61.9	174	107
HC	86.2	63	394
Total	77.0	281	457

Table 4: Classification matrix for discriminant analysis of temporal variations in the medical services

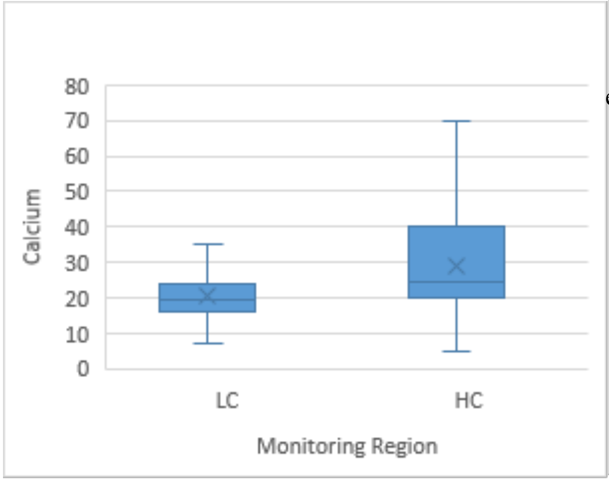
Monitoring Seasons	% Correct	Season Assigned by DA	
		Summer	Winter
Standard DA mode			
Summer	69.9	186	80
Winter	86.6	44	284
Total	79.1	266	328
Forward stepwise DA mode			
Summer	69.9	299	129
Winter	82.1	81	371
Total	76.1	428	452
Backward stepwise DA mode			
Summer	71.7	307	121
Winter	81.2	85	367
Total	76.6	428	452

Table 5: Classification functions (Eq. (1)) for discriminant analysis of spatial variation in the medical services

Parameters	Standard mode		Forward stepwise mode		Backward stepwise mode	
	LC coefficient	HC coefficient	LC coefficient	HC coefficient	LC coefficient	HC coefficient
Temperature	0.293	0.313				
EC	0.068	0.066				
TDS	-0.104	-0.099				
pH	66.557	65.851			49.024	48.652
Chloride	-0.382	-0.365			0.011	-0.004
Sodium	0.396	0.342	0.099	0.063	0.017	-0.003
Sulphate	-0.131	-0.139				
Calcium	0.63	0.811	1.319	1.675	0.51	0.667
Magnesium	-0.614	-0.823				
Potassium	2.556	2.388				
Total Hardness			-0.355	-0.42	-0.089	-0.101
Total Alkalinity	0.13	0.154	0.167	0.192	0.078	0.094
Bromide	15.182	14.267			-5.106	-5.874
Fluoride	-17.609	-18.323				
Nitrate	-1.2	-1.149				
Free Residual Chlorine	29.674	37.07	-0.896	6.447	15.602	21.068
Total Coliform	-4.546	-4.021				
Constant	-269.094	-267.334	-7.019	-10.126	-193.675	-192.702

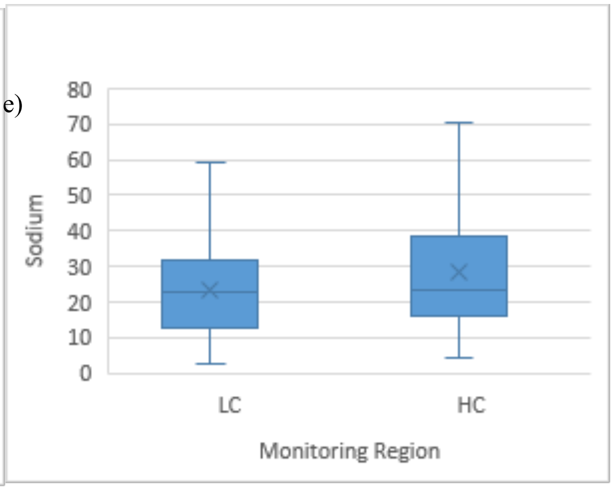
Table 6: Classification functions (Eq. (1)) for discriminant analysis of temporal variation in the medical services

Parameters	Standard mode		Forward stepwise mode		Backward stepwise mode	
	Summer coefficient	**Winter coefficient**	**Summer coefficient**	**Winter coefficient**	**Summer coefficient**	**Winter coefficient**
Temperature	1.422	1.078	1.381	1.038	1.155	0.872
EC	0.062	0.063				
TDS	-0.084	-0.089				
pH	69.848	68.655	65.164	63.955	54.884	53.964
Chloride	-0.44	-0.418				
Sodium	0.449	0.419				
Sulphate	-0.088	-0.103				
Calcium	0.739	0.75				
Magnesium	-1.26	-1.112				
Potassium	1.677	1.908	0.496	0.852	1.712	1.977
Total Hardness						
Total Alkalinity	0.238	0.21	0.311	0.284	0.194	0.177
Bromide	11.87	12.671				
Fluoride	-26.849	-24.164	-20.25	-17.129	-19.221	-16.727
Nitrate	-0.212	-0.506	-0.436	-0.753	0.433	0.167
Free Residual Chlorine	24.951	28.234	39.009	42.86	17.214	20.038
Total Coliform	-3.368	-3.603				
Constant	-304.059	-284.786	-282.665	-263.328	-237.474	-222.303

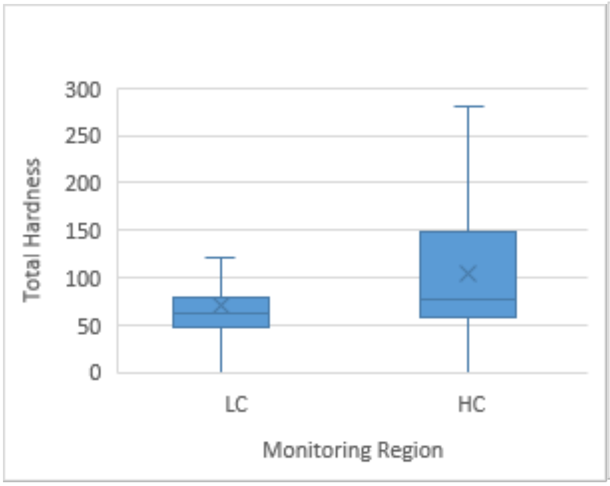


a)

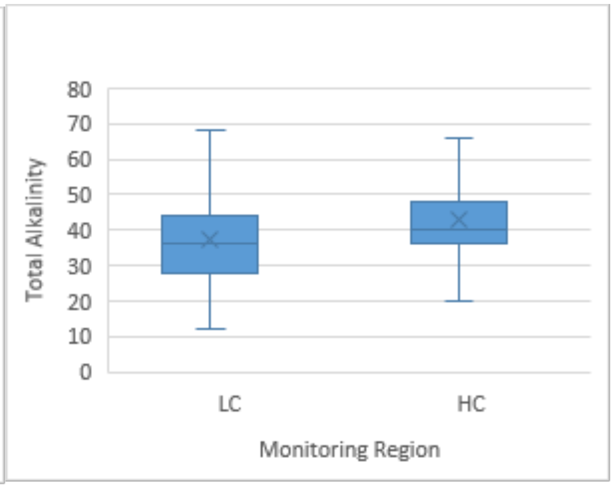
e)



b)



c)



d)

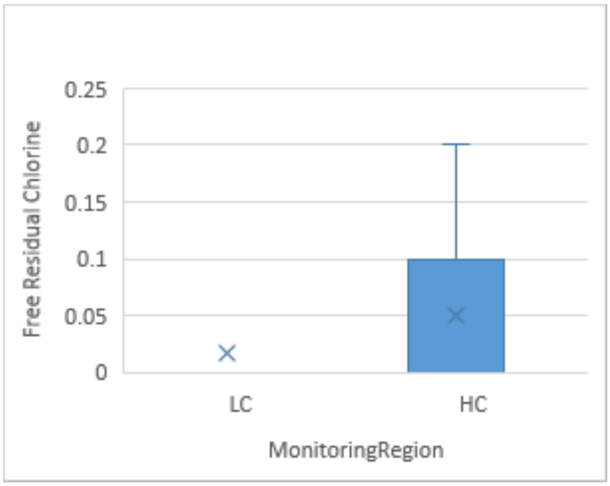


Figure 4: spatial variations: (a) Calcium; (b) Sodium; (c) Total Hardness; (d) Total Alkalinity; (e) Free Residual Chlorine in the medical services.

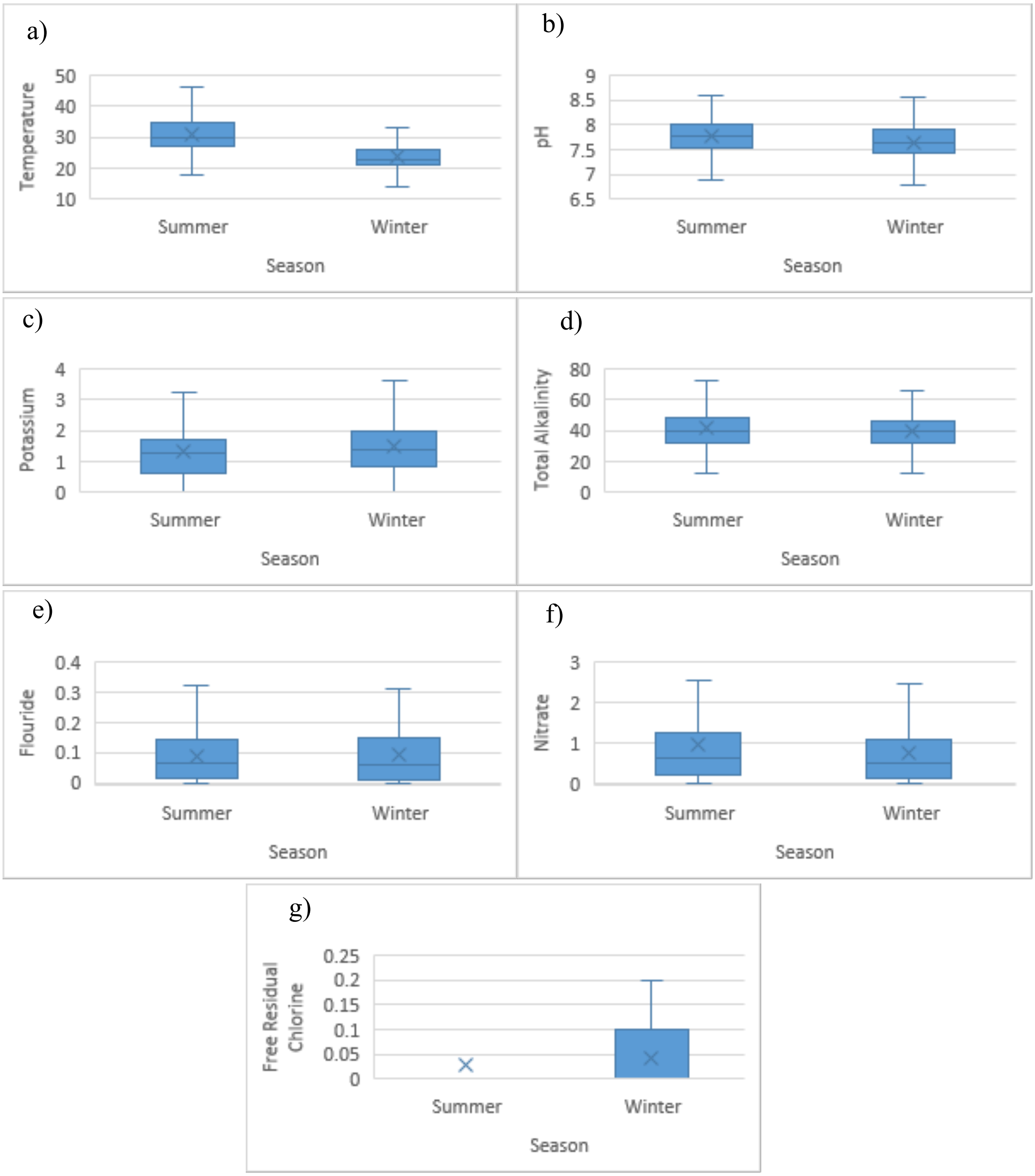


Figure 5: Temporal variations: (a) Temperature; (b) pH; (c) Potassium; (d) Total Alkalinity; (e) Fluoride (f) Nitrate (g) Free Residual Chlorine in the medical services.

4.4 Principle Component/Factor Analysis

Principal component/factor analysis was applied on the data sets (19 variables) separately for the two different categories, Spatial Variations (LC – HC) and Temporal Variations (Summer – Winter) as characterized by CA techniques, as the large set of variables is reduced into a smaller set of independent variables (principle components) containing the most essential information held in the larger set, in order to compare the compositional patterns among the analyzed drinking water samples and to determine the factors influencing each region involved in both categories. For the spatial variations, the input data resulted in the matrices (variables x cases) for PCA/FA giving [16 x 209] for LC and [17 x 380] for HC. As for temporal variations, winter was given in a matrix of [16 x 329] and [17 x 266] for summer. PCA of the two different data sets yielded five PCs for "LC" / "Summer", and three PCs for "HC" / "Winter" with Eigenvalues greater than 1, the factors with the highest Eigenvalues are the most significant. Results explained a total variance corresponding to water quality data sets for (LC – HC) are 74.84% and 64.86% and for (Summer & Winter) to be 76.64% and 67.1% respectively. Corresponding PCs, variable loadings and explained variance are presented in Table 7 and 8. Liu et al. (2003) classified the factor loadings as 'strong', 'moderate' and 'weak', corresponding to absolute loading values of >0.75, 0.75-0.50 and 0.50-0.30, respectively.

The data set referring to LC region, among five PCs, PC1, explaining 40.6% of total variance, has strong positive loading on [Total Hardness – EC – TDS – Cl – Na – Ca – Mg]. PC1 represents the impact of water desalination efficiency. The direct relationship between Total Hardness and Ca is natural as the presence of calcium in water increases the total hardness. Also, electrical conductivity is expected to increase with the increase of metal ions in water. PC2, explaining 10.95% of the total variance, has moderate positive loadings on [SO^{2-} – F – Nitrate], and moderate negative loadings on [Total Alkalinity]. PC2 represents the elements and compounds that gets released due to agriculture process and fertilizers. The inverse relationship between SO^{2-} and Total alkalinity is expected, as the acidic nature of the sulphate ion should have negative effect on total alkalinity. PC3 explaining 8.06% of the total variance, has moderate positive loadings on [Temperature – pH]. This factor can be interpreted as seasonal changes effect. PC4 explaining 7.87% of the total variance, has moderate positive loading on [Total Alkalinity]. PC5, explaining the lowest total variance 7.38%, has moderate positive loadings on [Free Residual Chlorine]. This factor represents the impact of post-treatment process and chlorine dosage.

Similarly, the data set referring to HC region, among three PCs, PC1, explaining 48.56% of total variance, has strong positive loading on [Total Hardness – EC – TDS – Cl – Na – Ca – Mg –SO^{2-} – Nitrite]. PC1 represents the impact of water desalination efficiency and agriculture process. Similar to the LC region, the direct relationship between Total Hardness and Ca is natural as the presence of calcium in water increases the total hardness. PC2, explaining 8.73% of the total variance, has moderate positive loadings on [Temperature – pH]. This factor represents seasonal changes impact. PC3 explaining 7.57% of the total variance, has moderate positive loadings on [K] and moderate negative loading on [F]. (Usman et. al (2014)) conducted a study in Terengganu, Malaysia, as the analysis was applied including 24 parameters regarding the spatial variations, resulting into three groups Less Polluted (LP), Moderately polluted (MP) and Highly Polluted (HP) in order to interpret probable impacts and compare the compositional patterns of the water samples that have been analyzed. The analysis resulted in eight major PCs with eigenvalues greater than 1. It was observed that the total variance of the original data structure had a value of 76.45%. All PCs held strong positive loadings such as Ca and Mg in PC1, Mn in PC3 and K in PC5. Similarly, (Y. Wang et al., 2013) applied the analysis based on spatial variations of the three groups LP, MP and HP, operating with 15 variables, setting the eigenvalue to be greater than 1, that resulted in five PCs holding 70.08%, 67.54% and 76.99% of the total variance in respective water quality datasets. Moreover, a case study was conducted by (Fan et al., 2010) in Pearl River Delta in China, with only seven parameters being included in the analysis and interpreting its spatial variations of the East River region, West River region and North River region, observing a total variance of 89.25%, 84.63% and 85.52% respectively. In this study, eigenvalue was selected based on the scree plot in order to determine the number of PCs to be used further in the analysis. The results have shown to give four PCs for the North and West River region. As for the North River region only PC1 with a total variance of 38.81% had one positive loading represented in the metal Hg. Likewise, for the West River region, only PC1 holding a 41.72% of the total variance contained positive loadings such as COD_{Mn}, TP and NH_3-N. As for the East River region, three PCs were obtained with all of them holding positive loadings.

As for the temporal variation, in the data referring to Summer, among five PCs, PC1, explaining 49.84% of total variance, has strong positive loading on [Total Hardness – EC – TDS – Cl – Na – Ca – Mg –SO^{2-}]. This factor also represents the impact of water desalination efficiency and agriculture effluents in water sources. PC2, explaining 7.89% of the total variance, has moderate positive loadings on [Free Residual Chlorine] and moderate negative loading on [K]. This factor explains the impact of chlorination process on drinking water. PC3 explaining 6.72% of the total variance, has strong positive loadings on [Temperature] and moderate negative loading on [pH]. PC3 indicates the inverse relationship between temperature and pH, which is a normal behavior. PC4 explaining 6.17% of the total variance, has moderate positive loading on [Total Coliform]. PC5, represent the lowest total variance 6.03%, has moderate positive loadings on [Total Alkalinity – Total Coliform]. PC4 and PC5 represent the impact of pathogenic contaminations.

Whereas in the Winter data set, among three PCs, PC1, explaining 48.26% of total variance, has strong positive loading on [Total Hardness – EC – TDS – Cl – Na – Ca – Mg –SO^{2-}]. As previous data sets, PC1 represents the impact of water desalination efficiency and agriculture impact. PC2, explains 10.11% of the total variance and has moderate positive loadings on [Fluoride], and strong negative loading on [Magnesium]. As for PC3, it explains 8.73% of the total variance, has moderate positive loadings on [Temperature – pH].

Table 7: Loadings of experimental variables (19) on significant principal components for LC and HC regions

Variables		PC1	PC2	PC3	PC4	PC5
LC Region (five significant principle components)						
Temperature		0.198	0.169	0.671	0.218	-0.163
EC		**0.833**	-0.228	0.251	-0.198	0.032
TDS		**0.839**	-0.147	0.274	-0.196	0.034
pH		0.077	0.39	0.609	0.178	-0.142
Chloride		**0.864**	-0.162	-0.022	-0.082	-0.219
Sodium		**0.892**	-0.054	-0.135	-0.129	-0.201
Sulphate		0.644	0.524	0.03	-0.171	0.033
Calcium		**0.801**	-0.21	0.079	0.206	0.294
Magnesium		**0.901**	-0.001	-0.123	-0.166	-0.049
Potassium		0.61	-0.112	-0.361	0.173	-0.457
Total Hardness		**0.899**	-0.15	0.011	0.086	0.19
Total Alkalinity		0.069	-0.584	0.052	0.675	0.261
Bromide		0.475	0.237	-0.218	0.391	0.332

Fluoride		0.291	0.619	-0.057	0.111	0.414
Nitrate		0.375	0.576	-0.314	0.381	-0.244
Free Residual Chlorine		0.087	0.052	-0.083	-0.419	0.546
Eigen Values		6.495	1.751	1.29	1.259	1.18
% Total Variance		40.595	10.946	8.06	7.868	7.376
Cumulative % Variance		40.595	51.541	59.601	67.468	74.844

HC Region (three significant principle components)						
Temperature		-0.063	0.627	0.29		
EC		**0.947**	-0.007	0.041		
TDS		**0.942**	0.046	0.048		
pH		0.057	0.611	0.172		
Chloride		**0.835**	-0.145	0.292		
Sodium		**0.896**	-0.224	0.169		
Sulphate		0.9	0.202	-0.084		
Calcium		**0.937**	0.01	-0.03		
Magnesium		**0.955**	-0.08	-0.054		
Potassium		0.401	-0.491	0.542		
Total Hardness		**0.96**	-0.022	-0.039		
Total Alkalinity		0.4	0.265	0.298		
Bromide		0.453	0.012	-0.449		
Fluoride		0.495	0.211	-0.565		
Nitrate		**0.757**	0.321	-0.055		
Free Residual Chlorine		0.301	-0.283	-0.338		
Total Coliform		0.079	0.24	0.154		
Eigen Values		8.255	1.484	1.287		
% Total Variance		48.56	8.73	7.571		
Cumulative % Variance		48.56	57.29	64.861		

Table 8: Loadings of experimental variables (19) on significant principal components for Summer and Winter seasons

Variables		**PC1**	**PC2**	**PC3**	**PC4**	**PC5**
Season - Summer (five significant principle components)						
Temperature		-0.019	0.078	**0.758**	-0.015	-0.152
EC		**0.916**	0.158	0.038	-0.12	-0.068
TDS		**0.911**	0.13	0.039	-0.058	-0.062
pH		0.205	0.324	-0.519	0.464	0.018
Chloride		**0.862**	-0.122	-0.108	-0.147	0.184
Sodium		**0.908**	-0.078	-0.136	-0.148	0.111
Sulphate		**0.893**	0.007	0.122	0.155	-0.107
Calcium		**0.928**	0.183	0.101	-0.069	0.016
Magnesium		**0.955**	0.069	-0.011	0.002	-0.069
Potassium		0.396	-0.607	-0.263	-0.378	0.223
Total Hardness		**0.953**	0.145	0.062	-0.045	-0.014
Total Alkalinity		0.294	0.402	0.196	-0.254	0.536
Bromide		0.632	-0.407	-0.032	0.251	-0.13
Fluoride		0.639	-0.148	0.103	0.324	-0.338
Nitrate		0.749	-0.278	0.09	0.117	0.008
Free Residual Chlorine		0.3	0.546	-0.254	-0.01	-0.066
Total Coliform		0.115	-0.124	0.211	0.592	0.675
Eigen Values		8.473	1.341	1.142	1.048	1.025
% Total Variance		49.841	7.888	6.717	6.166	6.027
Cumulative % Variance		49.841	57.728	64.445	70.611	76.638

Season - Winter (three significant principle components)						
Temperature		0.092	0.047	0.639		
EC		**0.94**	-0.077	-0.022		
TDS		**0.945**	-0.047	-0.001		
pH		-0.08	0.219	0.729		
Chloride		**0.759**	-0.436	0.237		
Sodium		**0.88**	-0.305	0.014		
Sulphate		**0.887**	0.271	-0.025		
Calcium		**0.915**	0.103	-0.157		
Magnesium		**0.949**	-0.004	-0.094		
Potassium		0.344	-0.768	0.278		
Total Hardness		**0.952**	0.068	-0.139		
Total Alkalinity		0.389	-0.137	-0.043		
Bromide		0.4	0.261	0.28		
Fluoride		0.415	0.582	0.166		
Nitrate		0.644	0.371	0.146		
Free Residual Chlorine		0.367	0.182	-0.374		
Eigen Values		7.722	1.617	1.396		
% Total Variance		48.262	10.108	8.725		
Cumulative % Variance		48.262	58.371	67.096		

5. CONCLUSION

Water quality monitoring programs present complex data sets; thus, in this study, in order to successfully assess the spatial and temporal variations in the water quality, several multivariate statistical techniques were applied to evaluate and analyze the drinking water quality parameters around different medical services in the State of Kuwait. The analysis of the collected data excluded 24 drinking water parameters among a total of 43 parameters due to their high missing percentage (above 50%). This was handled in order to avoid insignificant analysis results and the misrepresentation of the quality of drinking water. Hierarchical CA applied regarding the spatial variations classified the 14 sampling sites into two clusters, corresponding to Low Concentration (LC) and High Concentration (HC). It was observed that the distribution systems and water sources that supply the medical services in the southern part yield into lower concentrations than those in the northern part. Moreover, for the temporal variations, CA grouped the 12 months into two seasons, Summer and Winter, noticing a conflict between the months April and May, however, this result is acceptable as these two months are considered to be transition months. Cluster analysis was then followed and confirmed by both discriminant analysis and factor analysis. Regarding the DA, it interpreted a significant data reduction of the large data set, as only five (Sodium – Calcium – Total Hardness – Total Alkalinity – Chlorine) and seven (Temperature – pH – Potassium – Total Alkalinity – Fluoride – Nitrate – Chlorine) parameters were used concerning the spatial and temporal variations, respectively. These parameters are responsible for most of the variations present in the quality of drinking water. With reference to the PCA and FA, the analysis was conducted in order to extract correlations and resemblance among the parameters presented and to analyze and observe the drinking water samples in groups of similar quality. For the spatial variations, LC region identified five principle components that explained a total variance of 74.84%, as for the HC region three significant principle components were obtained explaining a total variance of 64.86%. Furthermore, concerning the temporal variations, summer season yielded into five principle components with a total variance of 70.6%, whereas the winter season resulted into three significant principle components describing a total variance of 67.1%. Thus, this demonstrates the practicality and suitability of the application of multivariate statistical techniques as a tool to analyze and interpret the complex water quality data sets provided, along with the understanding of their spatial and temporal variations, as well as adequate water quality assessment and management.

6. REFERENCES

Azhar, S. C., Aris, A. Z., Yusoff, M. K., Ramli, M. F., & Juahir, H. (2015). Classification of River Water Quality Using Multivariate Analysis. *Procedia Environmental Sciences*, *30*, 79–84. https://doi.org/10.1016/j.proenv.2015.10.014

Barakat, A., El Baghdadi, M., Rais, J., Aghezzaf, B., & Slassi, M. (2016). Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques. *International Soil and Water Conservation Research*, *4*(4), 284–292. https://doi.org/10.1016/j.iswcr.2016.11.002

Emmanuel, E., Pierre, M. G., & Perrodin, Y. (2009). Groundwater contamination by microbiological and chemical substances released from hospital wastewater: Health risk assessment for drinking water consumers. *Environment International*, *35*(4), 718–726. https://doi.org/10.1016/j.envint.2009.01.011

Fan, X., Cui, B., Zhao, H., Zhang, Z., & Zhang, H. (2010). Assessment of river water quality in Pearl River Delta using multivariate statistical techniques. *Procedia Environmental Sciences*, *2*, 1220–1234. https://doi.org/10.1016/j.proenv.2010.10.133

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J. M., & Fernandez, L. (2000). Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research*, *34*(3), 807–816. https://doi.org/10.1016/S0043-1354(99)00225-0

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/BF02291575

Kazi, T. G., Arain, M. B., Jamali, M. K., Jalbani, N., Afridi, H. I., Sarfraz, R. A., Baig, J. A., & Shah, A. Q. (2009). Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. *Ecotoxicology and Environmental Safety*, *72*(2), 301–309. https://doi.org/10.1016/j.ecoenv.2008.02.024

Liu, W.-C., Yu, H.-L., & Chung, C.-E. (2011). Assessment of Water Quality in a Subtropical Alpine Lake Using Multivariate Statistical Techniques and Geostatistical Mapping: A Case Study. *International Journal of Environmental Research and Public Health*, *8*(4), 1126–1140. https://doi.org/10.3390/ijerph8041126

Lovchinov, V., & Tsakovski, S. (2006). Multivariate statistical approaches as applied to environmental physics studies. *Open Physics*, *4*(2), 277–298. https://doi.org/10.2478/s11534-006-0012-3

Merkle, L. A., Layne, C. S., Bloomberg, J. J., & Zhang, J. J. (1998). Using factor analysis to identify neuromuscular synergies during treadmill walking. *Journal of Neuroscience Methods*, *82*(2), 207–214. https://doi.org/10.1016/S0165-0270(98)00054-5

Muangthong, S., & Shrestha, S. (2015). Assessment of surface water quality using multivariate statistical techniques: Case study of the Nampong River and Songkhram River, Thailand. *Environmental Monitoring and Assessment*, *187*(9), 548. https://doi.org/10.1007/s10661-015-4774-1

Muhammad, S., Shah, M. T., & Khan, S. (2011). Health risk assessment of heavy metals and their source apportionment in drinking water of Kohistan region, northern Pakistan. *Microchemical Journal*, *98*(2), 334–343. https://doi.org/10.1016/j.microc.2011.03.003

Ojo, O. A., & Adeniyi, I. F. (2012). The Impacts of Hospital Effluent Discharges on the Physico-chemical Water Quality of a Receiving Stream at Ile-Ife, Southwestern Nigeria. *Journal of Sustainable Development*, *5*(11), p82. https://doi.org/10.5539/jsd.v5n11p82

Ouyang, Y., Nkedi-Kizza, P., Wu, Q. T., Shinde, D., & Huang, C. H. (2006). Assessment of seasonal variations in surface water quality. *Water Research*, *40*(20), 3800–3810. https://doi.org/10.1016/j.watres.2006.08.030

Rapant, S., & Krčmová, K. (2007). Health risk assessment maps for arsenic groundwater content: Application of national geochemical databases. *Environmental Geochemistry and Health*, *29*(2), 131–141. https://doi.org/10.1007/s10653-006-9072-y

Sabatini, A. M. (2002). Identification of neuromuscular synergies in natural upper-arm movements. *Biological Cybernetics*, *86*(4), 253–262. https://doi.org/10.1007/s00422-001-0297-7

Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software*, *22*(4), 464–475. https://doi.org/10.1016/j.envsoft.2006.02.001

Singh, K. P., Malik, A., Mohan, D., & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—A case study. *Water Research*, *38*(18), 3980–3992. https://doi.org/10.1016/j.watres.2004.06.011

Singh, K. P., Malik, A., & Sinha, S. (2005). Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—A case study. *Analytica Chimica Acta*, *538*(1), 355–374. https://doi.org/10.1016/j.aca.2005.02.006

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed). Allyn and Bacon.

Usman, U. N., Toriman, M. E., Juahir, H., Abdullahi, M. G., Rabiu, A. A., & Isiyaka, H. (2014). Assessment of groundwater quality using multivariate statistical techniques in Terengganu. Science and Technology, 4(3), 42-49.

Velea, T., Gherghe, L., Predica, V., & Krebs, R. (2009). Heavy metal contamination in the vicinity of an industrial area near Bucharest. *Environmental Science and Pollution Research*, *16*(S1), 27–32. https://doi.org/10.1007/s11356-008-0073-5

Wang, X., Cai, Q., Ye, L., & Qu, X. (2012). Evaluation of spatial and temporal variation in stream water quality by multivariate statistical techniques: A case study of the Xiangxi River basin, China. *Quaternary International*, *282*, 137–144. https://doi.org/10.1016/j.quaint.2012.05.015

Wang, Y., Wang, P., Bai, Y., Tian, Z., Li, J., Shao, X., Mustavich, L. F., & Li, B.-L. (2013). Assessment of surface water quality via multivariate statistical techniques: A case study of the Songhua River Harbin region, China. *Journal of Hydro-Environment Research*, *7*(1), 30–40. https://doi.org/10.1016/j.jher.2012.10.003

Yu, P.-Y., Lin, Y. E., Lin, W.-R., Shih, H.-Y., Chuang, Y.-C., Ben, R.-J., Huang, W.-K., Chen, Y.-S., Liu, Y.-C., Chang, F.-Y., Yen, M.-Y., Liu, C.-C., Ko, W.-C., Lin, H.-H., & Shi, Z.-Y. (2008). The high prevalence of Legionella pneumophila contamination in hospital potable water systems in Taiwan: Implications for hospital infection control in Asia. *International Journal of Infectious Diseases*, *12*(4), 416–420. https://doi.org/10.1016/j.ijid.2007.11.008

Zhao, J., Fu, G., Lei, K., & Li, Y. (2011). Multivariate analysis of surface water quality in the Three Gorges area of China and implications for water management. *Journal of Environmental Sciences*, *23*(9), 1460–1471. https://doi.org/10.1016/S1001-0742(10)60599-2

Zhao, Y., Xia, X. H., Yang, Z. F., & Wang, F. (2012). Assessment of water quality in Baiyangdian Lake using multivariate statistical techniques. *Procedia Environmental Sciences*, *13*, 1213–1226. https://doi.org/10.1016/j.proenv.2012.01.115

Zhou, F., Liu, Y., & Guo, H. (2007). Application of Multivariate Statistical Methods to Water Quality Assessment of the Watercourses in Northwestern New Territories, Hong Kong. *Environmental Monitoring and Assessment*, *132*(1–3), 1–13. https://doi.org/10.1007/s10661-006-9497-x