SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ DERGİSİ

Sakarya University Journal of Science
SAUJS

ISSN 1301-4048 e-ISSN 2147-835X Period Bimonthly Founded 1997 Publisher Sakarya University
http://www.saujs.sakarya.edu.tr/

Title: A Case Study on the Relationship between Water Quality Parameters: Bursa

Authors: Ergun GUMUS

Recieved: 2022-03-05 00:00:00

Accepted: 2022-07-22 00:00:00

Article Type: Research Article

Volume: 26
Issue: 5
Month: October
Year: 2022
Pages: 1867-1878

How to cite
Ergun GUMUS; (2022), A Case Study on the Relationship between Water Quality Parameters: Bursa. Sakarya University Journal of Science, 26(5), 1867-1878, DOI: 10.16984/saufenbilder.1083427
Access link
http://www.saujs.sakarya.edu.tr/en/pub/issue/73051/1083427

Sakarya University Journal of Science 26(5), 1867-1878, 2022

A Case Study on the Relationship between Water Quality Parameters: Bursa

Ergun GUMUS*[1]

Abstract

Monitoring the quality of mains water in residential areas where industrialization is intense is of vital importance in terms of human health. For this purpose, quality parameters expressing the physical, chemical and biological properties of water are periodically observed through laboratory tests. During the evaluation of water quality, these parameters can be assessed individually or as a group by considering their interrelations. In this context, by using water quality reports of Bursa province which is an industrial city, answers to two questions were sought. The first of these questions is, getting evaluated on a group basis, which groups of water quality parameters are found to be highly correlated. The second question is whether the correlation between these interrelated parameter groups can be maintained in different measurement periods. For these purposes, analyzes were made using an approach which utilizes canonical correlation analysis, exhaustive scanning, and sliding window methods. As a result of these analyzes, it was observed that used approach gave successful results in terms of determining interrelated parameter groups and the differences in terms of interrelations between the measurement periods over these groups.

Keywords: Water quality parameters, canonical correlation analysis, exhaustive search, sliding window

1. INTRODUCTION

Water is the basic building block and indispensable element of all living organisms, from the smallest one to the largest. In terms of mass, 73% of our brain and heart, 79% of our skeletal muscles and kidneys, and 83% of our lungs are composed of water [1]. Water plays an important role in balancing our body temperature, digesting and transporting nutrients, excretion, working of our joints, protecting our brain and spinal cord against impacts, and many more. When all these aspects are taken into account, it becomes clear that human beings' access to water is of vital importance.

Water is essential not only for biological life, but also for agriculture, textile and industrial production. Water used in production is polluted with various chemicals and unfortunately is

* Corresponding author: ergun.gumus@btu.edu.tr
[1] Bursa Technical University,
ORCID: https://orcid.org/0000-0002-1327-6845

sometimes left back to the water cycle without being filtered. This situation especially affects drinking water quality of people living in cities that are intertwined with industry. For this reason, local governments frequently test the quality of the mains water in terms of physical, chemical and biological aspects. In these tests, (i) physical properties of water such as color, smell, taste, turbidity, (ii) chemical properties such as hardness, electrical conductivity, dissolved gases and elements in it, (iii) biological properties such as presence/number of viral pathogens and other microorganisms are examined. If the parameter values are within the standard ranges, the water source is considered to be suitable for use/drinking.

To date, many academic studies have been conducted on water quality parameters. In general, these studies focus on three main topics: Analysis of the relationship between quality parameters [2- 5], Depending on the analyzed relationship, estimation of a certain quality parameter [6- 9], and Estimation of total water quality using quality parameters [10- 14].

Among the studies on analysis of relationships between quality parameters, Noori et al. [2] examined physical and chemical parameters obtained from the Karoon River using Canonical Correlation Analysis. As a result of the research, it was reported that parameters of electrical conductivity, total dissolved solids, total amount of ions and water hardness have a high weight in explaining the canonical diversity. In a similar study [3], measurements taken from Macau main storage reservoir for ten years were used. Principal Component Analysis, a dimension reduction tool, was applied on two different groups of physical and chemical parameters. The canonical relations between remaining parameters were examined and it was seen that electrical conductivity from the physical parameters group and amount of chloride from the chemical parameters group had the highest canonical weight in their own group. In another study, Parmar et al. [4] divided water quality measurements into sub-levels using Discrete Wavelet Transform method instead of using them directly like in current studies. They obtained measurement-signal vectors of these levels. Afterwards, they showed that the chemical oxygen demand parameter was highly correlated with other three chemical parameters. On the other hand, Sallam et al. [5], established regression models for estimation of parameters such as pH, dissolved oxygen, electrical conductivity, and total dissolved solids using air temperature, relative humidity, and quantitative/qualitative properties of the water released into the water source. They stated that there is a strong correlation between these parameters.

This study focused on the relationship between water quality parameters. In this context, six-years-long water quality measurements of Bursa province were used and two separate subjects were examined. First subject is to identify two subgroups of quality parameters that provide a high canonical correlation. For this purpose, Exhaustive Search (ES) method was used. The second subject of the study is to determine whether there is a significant correlation between two different measurement periods or not. At this point, a sliding window approach along with Canonical Correlation Analysis (CCA) was applied on the data which is in form of time series.

Rest of the study is organized as follows: In the second part, the dataset and approaches used are introduced. In the third part, experimental results are given. The fourth and final part is devoted to conclusions.

2. MATERIAL AND METHODS

2.1. The Dataset

In the study, quality of drinking water reports published weekly by Bursa Water and Sewerage Administration were used (www.buski.gov.tr/SuAnalizRaporu/Detay/). There are 257 measurement reports in total for the period between June 15th, 2014 and May 6th, 2020 (70 months). Distribution of these reports by months and years can be seen in Table 1.

Table 1 Number of measurements on the basis of Months&Years

	Month												
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total**
2014						1	1	1	1	1	2	1	8
2015	1	1	1	1	1	1	5	4	4	4	4	5	32
2016	2	4	5	4	2	5	3	5	3	4	5	4	46
2017	4	3	5	4	5	4	4	4	4	4	5	4	50
2018	5	4	4	4	5	4	4	4	4	5	4	4	51
2019	5	4	4	4	5	3	5	4	4	5	4	4	51
2020	5	4	4	5	1								19

In the aforementioned reports, there are a total of 54 parameters regarding the physical, chemical and biological properties of drinking water. For some of these parameters, data before a specific date is not available. At the same time, for some parameters, number ranges are used instead of exact values. There are also parameters that have the same value for each measurement or that can be neglected because their standard deviation is very low. For these reasons, 18 parameters which are shown in Table 2 were selected to be used in the study.

Table 2 Water quality parameters used in the study

Par#	Name of the parameter	[Min - Max]	Mean	Standard Deviation
1	pH	[7.36 – 8.41]	7.889	0.260
2	Blurriness	[0.02 – 0.91]	0.264	0.131
3	Total Hardness	[137.8 – 250.9]	177.463	26.331
4	Total Iron	[0 - 64]	21.659	10.490
5	Manganese	[1 - 30]	12.284	4.147
6	Chloride	[3.8 – 20.37]	8.662	2.666
7	Sulfate	[7 - 51]	19.548	11.283
8	Nitrate	[0 – 5.3]	0.991	0.702
9	Conductivity	[277 - 432]	356.813	36.961
10	Permanganate Index	[0.12 – 4.06]	0.996	0.406
11	Sodium	[3.94 – 13.255]	7.787	1.880
12	Aluminum	[0 - 163]	32.458	29.314
13	Fluoride	[0 – 0.17]	0.065	0.021
14	Free Chlorine	[0.42 – 0.78]	0.648	0.062
15	Arsenic	[0.017 – 5.914]	2.637	1.271
16	Copper	[0 – 6.57]	0.710	0.725
17	Nickel	[0.25 – 9.22]	3.201	1.256
18	Trihalomethane (THM)	[0.42 - 64]	30.117	8.981

There is a missing data problem for some of the selected parameters. To overcome this problem, the "window average" approach was used, and each missing data area was filled with the average of the other data in its 14 neighborhood. In the last step, all parameters were normalized to the range [0,1] since they take values from different number ranges.

Measurement reports were fetched from website of Bursa Water and Sewerage Administration using Python library “Beautiful Soup”. All analyzes were carried on PC platform using MATLAB.

2.2. Canonical Correlation Analysis

Canonical Correlation Analysis (CCA) method is a statistical analysis tool which is used to find canonical projection vectors that transfer samples of two different observation sets (views) belonging to the same entity to ideal data spaces in order to maximize their correlation (Pearson's correlation) between [15]. The method is used in many research areas like feature selection in classification processes [16, 17], biomedical applications [18- 20], bioinformatics [21, 22], and so on.

CCA can be briefly stated as follows:

Suppose that we have two zero-mean views $S_X = [x_1, x_2, \ldots, x_N]$, and $S_Y = [y_1, y_2, \ldots, y_N]$ in *D*-dimensional space for *N* samples. We can project these samples into a one-dimensional space using two basis vectors w_X and w_Y, and write the correlation coefficient *r* between these projections as seen in Eq. 1.

$$r = \frac{w_X{}^T S_X S_Y{}^T w_Y}{\sqrt{(w_X{}^T S_X S_X{}^T w_X)(w_Y{}^T S_Y S_Y{}^T w_Y)}} = \frac{w_X{}^T C_{XY} w_Y}{\sqrt{(w_X{}^T C_{XX} w_X)(w_Y{}^T C_{YY} w_Y)}} \quad (1)$$

Here, C_{XX} and C_{YY} are within-class covariance matrices and $C_{XY} = C_{YX}{}^T$ are between-class covariance matrices. Our goal is to find the ideal canonical projection vectors w_X and w_Y that will maximize coefficient *r*. For this, we can create the optimization problem seen in Eq. 2 by specifying a set of constraints and using the Lagrangian relaxation method.

$$L(\lambda, w_X, w_Y) = w_X{}^T C_{XY} w_Y - \frac{\lambda}{2}(w_X{}^T C_{XX} w_X - 1) - \frac{\lambda}{2}(w_Y{}^T C_{YY} w_Y - 1) \quad (2)$$

By solving our optimization problem, the equality seen in Eq. 3 and the eigenproblem seen in Eq. 4 are obtained. The coefficient λ is the Lagrange multiplier from our optimization problem and is also the square root of the eigenvalue of our eigenproblem.

$$w_Y = \frac{C_{YY}{}^{-1} C_{YX} w_X}{\lambda} \quad (3)$$

$$C_{XX}{}^{-1} C_{XY} C_{YY}{}^{-1} C_{YX} w_X = \lambda^2 w_X \quad (4)$$

With the solution of our eigenproblem, the canonical projection vector w_X (eigenvector) and the coefficient λ (eigenvalue) are obtained. By substituting both of them in Eq. 3, the canonical projection vector w_Y is obtained.

2.3. Exhaustive Search

In the first stage of the study, it was aimed to find subgroups of water quality parameters that have a significant relationship. Some of previous studies [2, 3] used CCA method to examine the relationship between physical and chemical quality parameters for this purpose. Unlike their work, instead of using views of two logical classes, relationships between mixed subsets (views) of all parameters given in Table 2 were examined in this study. In literature, there are iterative [23] and adaptive [24] techniques for composing CCA views. However, since the number of parameters subject to the study is appropriate, Exhaustive Search (ES) approach, in which all possible subsets of parameters are used, was applied. The algorithm regarding to application of the approach together with CCA can be seen in Figure 1.

Like many machine learning methods, CCA can be affected by the way training and test sets are formed, and it can be under the influence of bias effect. For this reason, the algorithm was run using 5×2 cross validation, and 10 different test scenarios were obtained.

```
N ← number of samples
D ← number of parameters
Par ← {1, 2, …, D}
Data ← N × D dimensional data matrix
train ← indices of training samples
test ← indices of test samples

for i ← 1 to 2^D – 1 do
    f ← i^th subset of Par                                              / Except null set
    if number_of_elements(f) > 1 then
        half ← ⌊ number_of_elements(f) / 2⌋
        for k ← 1 to half do
            x ← a k-element combination of f
            y ← f – x
            train_view1 ← Data(train, x)
            train_view2 ← Data(train, y)
            test_view1 ← Data (test, x)
            test_view2 ← Data (test, y)
            [w_x, w_y] ← CCA(train_view1, train_view2)                  / CCA training
            r ← pearson_correlation( test_view1 × w_x^T , test_view2 × w_y^T )  / CCA test
        end for
    end if
end for
```

Figure 1 Implementation of CCA using ES

2.4. Sliding Window Approach

Another subject of study is to find out whether there is a significant correlation between water quality parameters in different measurement periods when measurements are ordered from oldest to newest according to their dates. For this purpose, the sliding window approach seen in Figure 2 was used. Accordingly, two observation periods (Period1 and Period2), which contain equal number of observations (W) but do not overlap with each other were created. Then, canonical projection vectors which maximize CCA training correlation of two views in Period1 were calculated. Lastly, using same canonical projection vectors, samples in both views of Period2 were projected to a new data space and test correlation of projected views, which resolves a possible relationship between two observation periods, was obtained. In each iteration of the approach, one of the observation periods (Period1) remained constant while the other (Period2) was shifted by one sample forward. In case where Period2 could not be shifted any further, the approach was repeated from the beginning by shifting Period1 one sample further. Whole process is maintained until the case where two periods overlap.

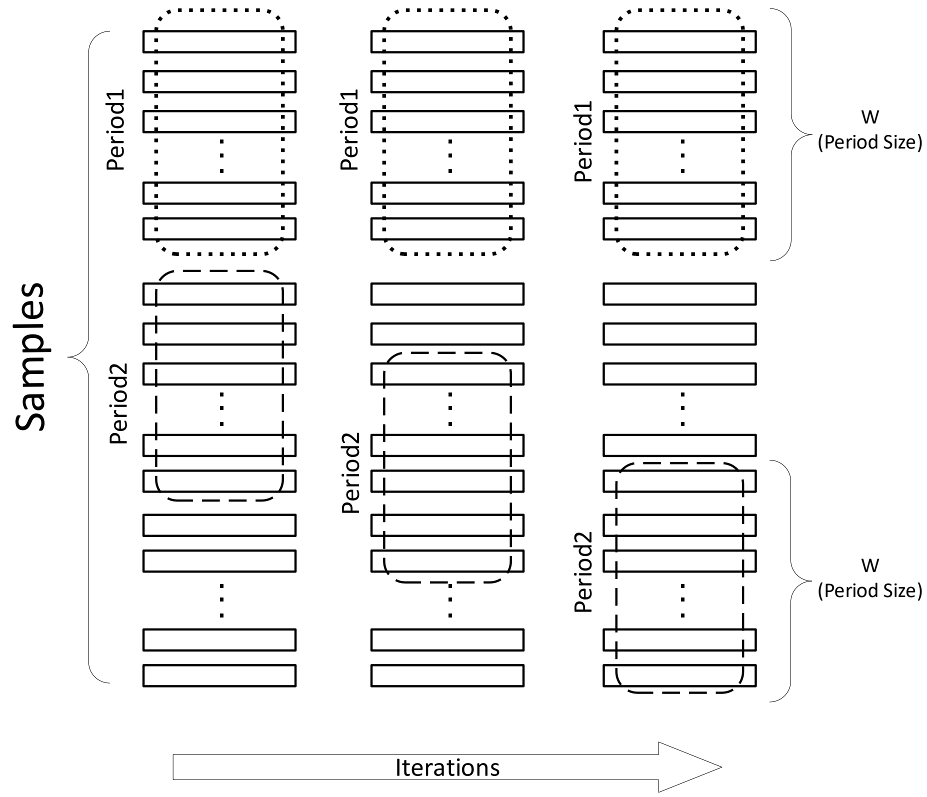

Figure 2 Sliding window approach for detecting periodic relationships

3. RESULTS

3.1. Group-Based Relationships among Water Quality Parameters

In order to find an answer to the question of whether there is a significant relationship between water quality parameters on a group basis, the algorithm mentioned in Figure 1 was applied to whole dataset, with a 5×2 cross-validation approach. In this way, 10 different test scenarios were created, each containing $2^{18} - 19$ test correlations. The 5-number summary obtained after eliminating extreme values for each of these scenarios is shown in Table 3.

Table 3 5-number summary of test scenarios

Test #	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
1	0.3938	0.6966	0.8064	0.8985	0.9510
2	0.3056	0.6520	0.7935	0.8829	0.9282
3	0.3987	0.6982	0.8386	0.8978	0.9592
4	0.3081	0.6543	0.7815	0.8852	0.9340
5	0.4205	0.7041	0.7991	0.8931	0.9409
6	0.3202	0.6596	0.7969	0.8859	0.9540
7	0.3419	0.6678	0.8037	0.8852	0.9322
8	0.3549	0.6791	0.7849	0.8952	0.9557
9	0.3644	0.6720	0.8145	0.8772	0.9404
10	0.3114	0.6637	0.8024	0.8985	0.9424

When Table 3 is examined, it is seen that there is not a big difference between upper quartile values of the test scenarios. The same is also true for maximum values. At this point, top 100 test correlations for each scenario and corresponding water quality parameter groups (View1 parameters and View2 parameters) were detected. Afterwards, among these 100×10 cases, the most common parameter group in View1 was scanned. As a condition for this scan, selection of parameter groups in View1 which took place in at least half of cases was taken as a basis. Result of this scan is given in Table 4.

Table 4 Most frequently encountered parameter groups in selected test scenarios (View1)

Parameter Group	Observation Rate (%)
{Total Hardness, Chloride}	86.2
{Total Hardness, Chloride, Permanganate Index}	70.4
{Total Hardness, Chloride, Arsenic}	55.6

For View1, it was observed that high test correlations had been obtained by using three parameter groups formed by quality parameters #3 (Total Hardness), #6 (Chloride), #10 (Permanganate Index), and #15 (Arsenic).

Most frequently observed parameter group for View1 is the group consisting of “Total Hardness” and “Chloride” pair. This group took place in a total of 862 cases. Among all cases, most frequently observed parameter groups for View2 were scanned with a similar approach, where selection of parameter groups in View2 which took place in at least half of cases was taken as a basis. Result of this scan is given in Table 5.

Table 5 Most frequently encountered parameter groups in selected test scenarios (View2)

Parameter Group	Observation Rate (%)
{Conductivity, Sodium}	100
{Nitrate, Conductivity, Sodium}	96.9
{Nitrate, Conductivity, Sodium, Fluoride}	77.8
{Sulphate, Nitrate, Conductivity, Sodium}	77.3
{Nitrate, Conductivity, Sodium, THM}	62

For View2, it was observed that high test correlations had been obtained by using five parameter groups formed by quality parameters #7 (Sulphate), #8 (Nitrate), #9 (Conductivity), #11 (Sodium), #13 (Fluoride), and #18 (THM). “Conductivity” and “Sodium” parameters were encountered in all of these five groups.

After application of cross-validation approach, CCA test correlations ranging from 0.9021 to 0.9434 were obtained between View1 created with “Total Hardness” and “Chloride” parameters and View2 created with “Conductivity” and “Sodium” parameters.

Considering average test correlations obtained from 10 test scenarios which were formed by 5×2 cross-validation approach, the highest average test correlation was 0.9402. This value was obtained by using {Blurriness, Total Hardness, Chloride, Permanganate Index} parameters for View1, and {Manganese, Nitrate, Conductivity, Sodium, Fluoride, THM} parameters for View2.

3.2. Relationship between Observation Periods

For the secondary purpose of the study, observations, which were ordered from oldest to newest according to measurement dates, were converted into periods covering 12 months. After that, couples of periods were analyzed using the sliding window approach mentioned in Section 2.4 in order to check for existence of a similarity/difference between them. In this context, as seen in Table 1, observations covering 6 years were divided into consecutive 12-months-long periods with a one-month progress, and a total of 61 periods were obtained. While forming these periods, if there were more than one observation belonging to some specific month, only one of these observations was randomly selected. Since this selection is a stochastic process, each pair of non-overlapping periods (Period1 and Period2) whose relationship will be questioned were created 1000 times and averages of training and test correlations were used in analysis. While calculating both training and test correlations, {Blurriness, Total Hardness, Chloride, Permanganate Index} parameters for View1 and {Manganese, Nitrate, Conductivity, Sodium, Fluoride, THM} parameters for View2 were used in each period. The reason for this choice is that the highest possible test correlation was obtained with the relationship between these parameter groups.

In trials using the sliding window approach, first 49 of 61 periods were used to calculate i) training correlations between views of these periods, and ii) canonical projection vectors that maximize these correlations. Out of 1000 trials, training (Period1) correlations obtained using these periods are shown in Figure 3 together with outliers in box graphic form.

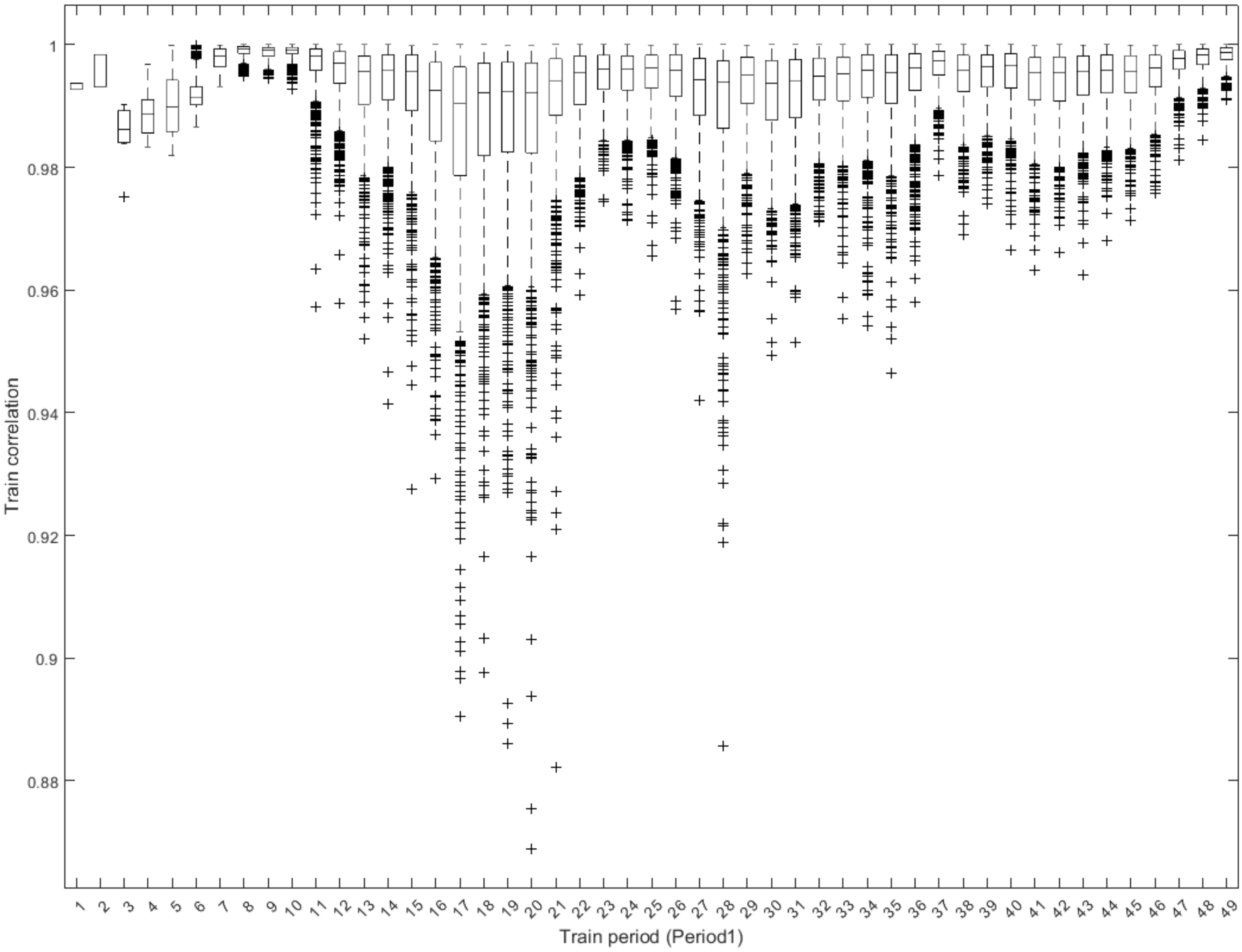

Figure 3 Training correlations obtained during various train periods

When Figure 3 is examined, it is seen that high training correlations ranging from 0.87 to 0.99 can be obtained in each of the different train periods, as would be expected from CCA approach. It is also seen that correlation values obtained at the end of multiple trials per period do not show much variation for train periods including initial months with few observations. Lowest outliers (+ symbol) in training correlation are seen in train periods #17, #18, #19, #20, #21, and #28 (5 consecutive 12-months-long periods covering October 2015 – January 2017 interval, and a single 12-months-long period covering September 2016 – August 2017 interval).

In accordance with the sliding window approach seen in Figure 2, test correlation between views of Period2 (test period) were obtained by applying canonical projection vectors which maximize the training correlation between views of Period1 (train period). Accordingly, *i.* period was accepted as Period1 and each of the following periods in the interval [*i* + 12, 61] was considered as Period2 one by one. In this way, the relationship of each Period1 with each subsequent Period2 was evaluated through the test correlation obtained. The boxplot created using test correlations obtained over 1000 trials is shown in Figure 4.

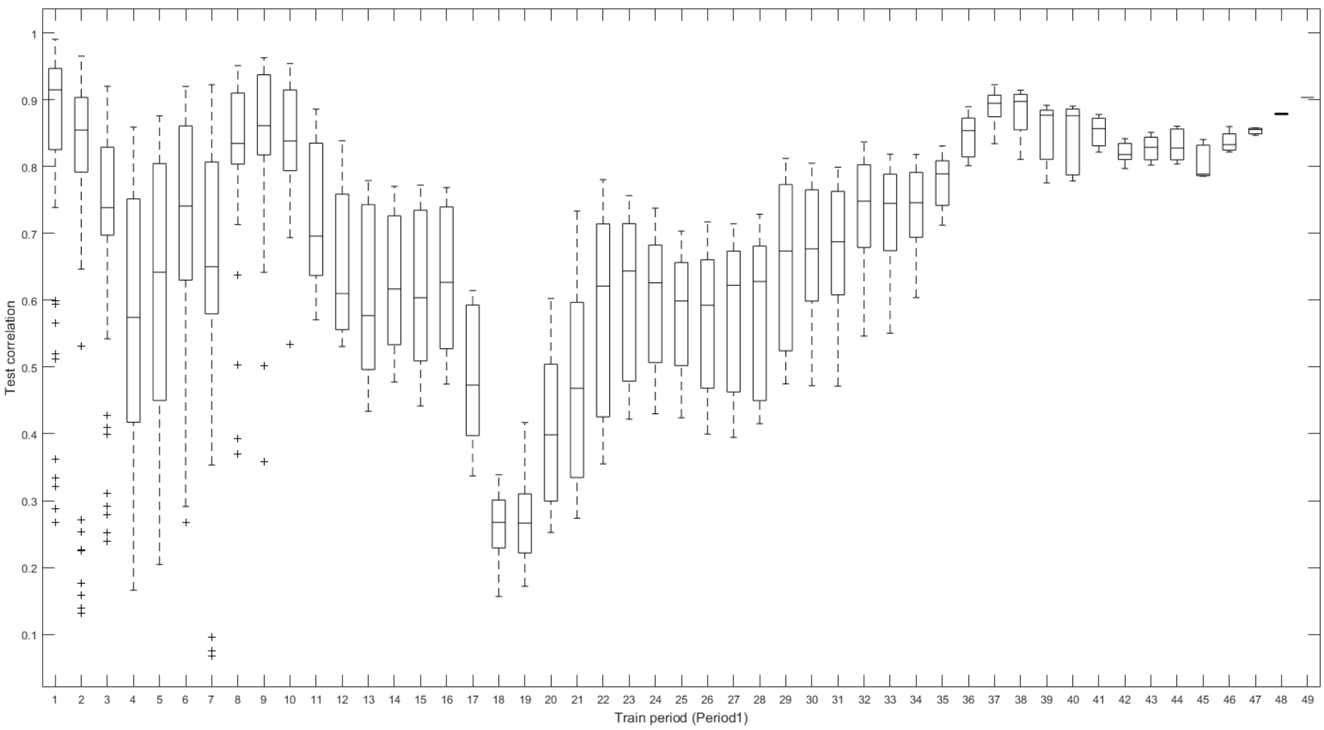

Figure 4 Test correlations obtained by applying canonical vectors of training period (Period1) to views of test period (Period2)

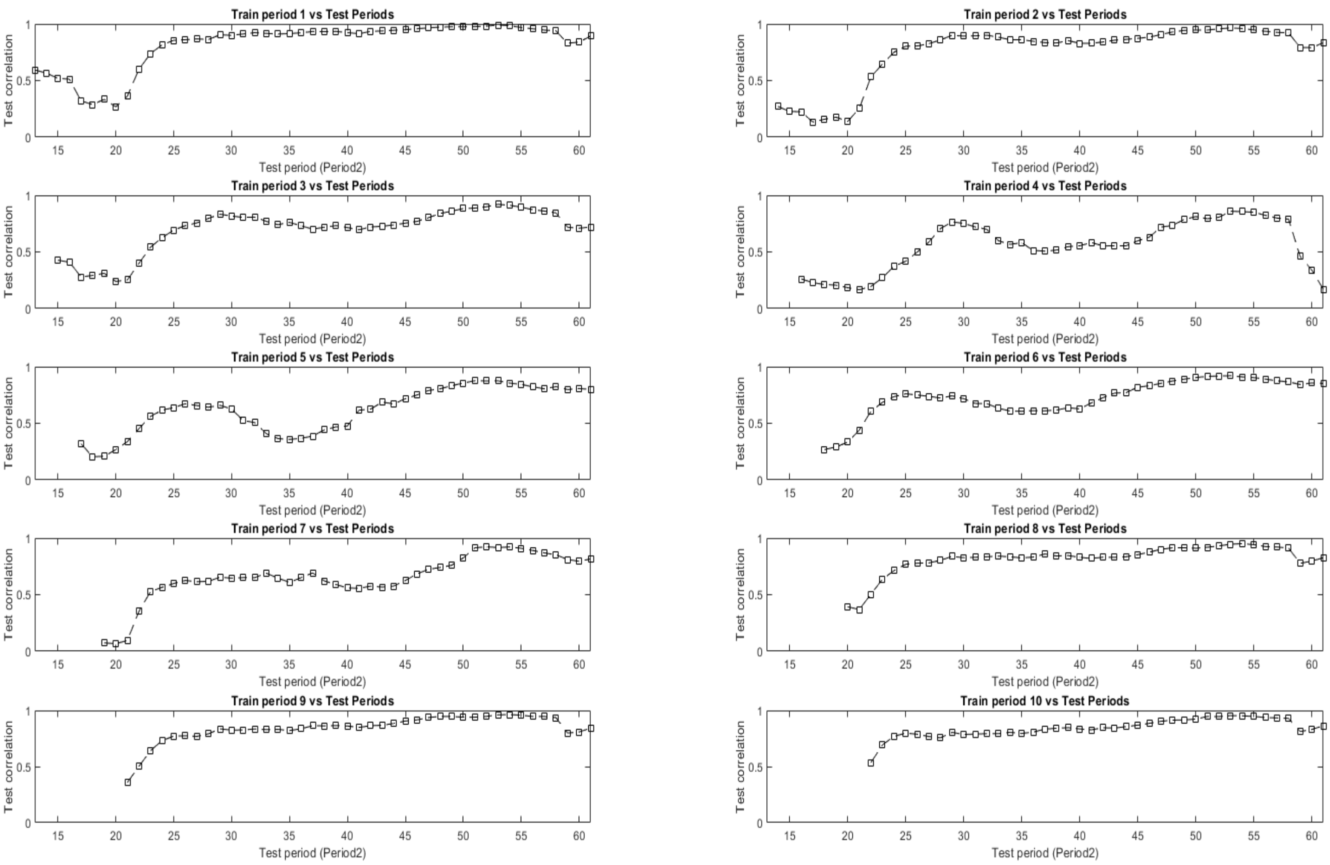

Figure 5 Variation of mean test correlations obtained using first 10 training periods

As can be seen in Figure 4, median values of test correlations obtained by applying canonical vectors of training period (Period1) to views of test period (Period2) remained low (at the level of 0.2665) especially for 18th and 19th training periods (two consecutive periods of 12 months each, covering the range of November 2015 - November 2016). For 18th training period, mean and standard deviation of test correlations were as low as 0.2603 and 0.0507, respectively. Similarly, for 19th training period, mean and standard deviation of test correlations were as low as 0.2722 and 0.0666, respectively. Median training correlations obtained for these two periods were similar, about 0.99. Other than this, test correlations which can be defined as outliers (+ symbols) were observed using first 10 training periods. Variation of mean test correlations for these periods is shown in Figure 5.

Examining Figure 5, it is clear that mean test correlation values remain low until 20th test period. Only after that, a rise is possible. One might suspect that, the reason is lack of diversity in observations for leading periods (check number of measurements for leading months in Table 1). However, although observation diversity is low for first 10 periods, it is seen that canonical vectors obtained in training phases of these periods provide high mean test correlations after 20th test period. Moreover, checking Table 1, we can see that lack of diversity in observations comes to an end after 13th period.

4. CONCLUSIONS

In this study, a case analysis was conducted i) to determine the relationship between water quality parameter groups, and ii) to determine whether there exists a periodic relationship between measurements taken at different times or not. In this context, exhaustive search and sliding window approaches are employed along with canonical correlation analysis. When the relationship of quality parameters on a group basis is evaluated, it is seen that the relation between the group consisting of {Blurriness, Total Hardness, Chloride, Permanganate Index} parameters and the group consisting of {Manganese, Nitrate, Conductivity, Sodium, Fluoride, THM} parameters is maximized using the sets of observations in accordance with 5×2 cross-validation. In addition, by using detected groups of quality parameters, the relationship between distinct 12-months-long periods was also questioned. As a result of this inquiry, it was seen that the relationship established between these two parameter groups was valid in most of the observation periods covering 70 months. However, especially for the periods covering November 2015 - November 2016 interval, low-level test correlations such as 0.2603 and 0.2722 were observed, which means the relationship established between detected parameter groups in this interval does not hold in other observation periods. This gives us an idea about benefit of applied methodology in order to detect seasonal variability of the relationship between water quality parameters.

Acknowledgments
The author would like to thank reviewers and editors for their valuable time.

Funding
The author has not received any financial support for the research, authorship or publication of this study.

The Declaration of Conflict of Interest/ Common Interest
No conflict of interest or common interest has been declared by the author.

Authors' Contribution
The author contributed 100% to the study.

The Declaration of Ethics Committee Approval
This study does not require ethics committee permission or any special permission.

The Declaration of Research and Publication Ethics
The author of the paper declare that he complies with the scientific, ethical and quotation rules of SAUJS in all processes of the paper and that he

does not make any falsification on the data collected. In addition, he declares that Sakarya University Journal of Science and its editorial board have no responsibility for any ethical violations that may be encountered, and that this study has not been evaluated in any academic publication environment other than Sakarya University Journal of Science.

REFERENCES

[1] H. H. Mitchell, T. S. Hamilton, F. R. Steggerda, H. W. Bean, "The chemical composition of the adult human body and its bearing on the biochemistry of growth," in The Journal of Biological Chemistry, 158, pp. 625-637, 1945.

[2] R. Noori, M. S. Sabahi, A. R. Karbassi, A. Baghvand, H. Taati Zadeh, "Multivariate statistical analysis of surface water quality based on correlations and variations in the data set," in Desalination, 260, pp. 129-136, 2010.

[3] M. C. Chan, I. Lou, W. K. Ung, K. M. Mok, "Integrating principle component analysis and canonical correlation analysis for monitoring water quality in storage reservoir," in Applied Mechanics and Materials, 284-287, pp. 1458-1462, 2013.

[4] K. S. Parmar, R. Bhardwaj, "Wavelet and statistical analysis of river water quality parameters," in Applied Mathematics and Computation, 219, pp. 10172-10182, 2013.

[5] G. A. H. Sallam, E. A. Elsayed, "Estimating relations between temperature, relative humidity as independent variables and selected water quality parameters in Lake Manzala, Egypt," in Ain Shams Engineering Journal, 9, pp. 1-14, 2018.

[6] E. Dogan, B. Sengorur, R. Koklu, "Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique," in Journal of Environmental Management, 90, pp. 1229-1235, 2009.

[7] M. J. Alizadeh, M. R. Kavianpour, "Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean," in Marine Pollution Bulletin, 98, pp. 171-178, 2015.

[8] I. Seo, S. H. Yun, S. Y. Choi, "Forecasting water quality parameters by ANN model using preprocessing technique at the downstream of Cheongpyeong dam," in Procedia Engineering, 154, pp. 1110-1115, 2016.

[9] S. Mazhar, A. Ditta, L. Bulgariu, I. Ahmad, M. Ahmed, A. A. Nadiri, "Sequential treatment of paper and pulp industrial wastewater: Prediction of water quality parameters by Mamdani fuzzy logic model and phytotoxicity assessment," in Chemosphere, 227, pp. 256-268, 2019.

[10] G. A. Cordoba, L. Tuhovcak, M. Taus, "Using artificial neural network models to assess water quality in water distribution networks," in Procedia Engineering, 70, pp. 399-408, 2014.

[11] A. D. Sutadian, N. Muttil, A. G. Yilmaz, B. J. C. Perera, "Using the analytic hierarchy process to identify parameter weights for developing a water quality index," in Ecological Indicators, 75, pp. 220-233, 2017.

[12] G. Sotomayor, H. Hampel, R. F. Vazquez, "Water quality assessment with emphasis in parameter optimisation using pattern recognition methods and genetic algorithm," in Water Research, 130, pp. 353-362, 2018.

[13] A. N. Ahmed, F. B. Othman, H. A. Afan, R. K. Ibrahim, C. M. Fai, M. S. Hossain, M. Ehteram, A. Elshafie, "Machine learning methods for better water quality prediction," in Journal of Hydrology, 578, 124084, 2019.

[14] M. Tripathi, S .K. Singal, "Use of principal component analysis for parameter selection for development of a novel water quality index: A case study of river Ganga India," in Ecological Indicators, 96, pp. 430-436, 2019.

[15] D. R. Hardoon, S. Szedmak, J. S. Taylor, “Canonical Correlation Analysis: An overview with application to learning methods,” in Neural Computation, 16(12), pp. 2639-2664, 2004.

[16] C. O. Sakar, O. Kursun, F. Gurgen, “A feature selection method based on kernel canonical correlation analysis and the minimum Redundancy-Maximum Relevance filter method,” in Expert Systems with Applications, 39(3), pp. 3432-3437, 2012.

[17] W. Yan, C. Shuang, Y. Hongnian, “Mutual information inspired feature selection using kernel canonical correlation analysis,” in Expert Systems with Applications: X, 4, 100014, 2019.

[18] D. Lin, V. D. Calhoun, Y. Wang, “Correspondence between fMRI and SNP data by group sparse canonical correlation analysis,” in Medical Image Analysis, 18(6), pp. 891-902, 2014.

[19] W. Xingjie, Z. Ling-Li, S. Hui, L. Ming, H. Yun-an, H. Dewen, “Blind source separation of functional MRI scans of the human brain based on canonical correlation analysis,” in Neurocomputing, 269, pp. 220-225, 2017.

[20] A. S. Janani, T. S. Grummett, T. W. Lewis, S. P. Fitzgibbon, E. M. Whitham, D. DelosAngeles, H. Bakhshayesh, J. O. Willoughby, K. J. Pope, “Improved artefact removal from EEG using Canonical Correlation Analysis and spectral slope,” in Journal of Neuroscience Methods, 298, pp. 1-15, 2018.

[21] M. G. Naylor, X. Lin, S. T. Weiss, B. A. Raby, C. Lange, “Using canonical correlation analysis to discover genetic regulatory variants,” in PLoS ONE, 5(5), e10395, 2010.

[22] Y. Zhang, J. Zhang, Z. Liu, Y. Liu, S. Tuo, “A network-based approach to identify disease-associated gene modules through integrating DNA methylation and gene expression,” in Biochemical and Biophysical Research Communications, 465(3), pp. 437-442, 2015.

[23] L. Liu, Q. Wang, E. Adeli, L. Zhang, H. Zhang, D. Shen, “Feature selection based on iterative canonical correlation analysis for automatic diagnosis of Parkinson’s disease,” in International Conference on Medical Image Computing and Computer-Assisted Intervention, 9901, pp. 1-8, 2016.

[24] W. Hu, D. Lin, S. Cao, J. Liu, J. Chen, V.D. Calhoun, Y. Wang, “Adaptive sparse multiple canonical correlation analysis with application to imaging (epi)genomics study of schizophrenia,” in IEEE Transactions on Biomedical Engineering, 65(2), pp. 390-399, 2019.