Reliability of Water Resources Quality Monitoring Program Data

Sabah Obaid Hamad Al-Shujairi
Environmental Department of Central Region-Ministry of Environment- Iraq
E-mail: dr.sabah2012@gmail.com

Abstract

Data are a vital component of water resources management activities. The consequences of using poor quality data include faulty decisions, higher risk to the environment or human health, wasted resources and loss of credibility. Box-whisker plot, cations-anions balance and relative total dissolved solids (TDS) to electrical conductivity (EC) ratio were used to examine validity of water quality data. These techniques and reliability check applied on water quality data of Tigris river in 2013 to ensure that the data can be used for decision making in the management of water resources with a high level of confidence, improve the reliability of water quality assessments, and discovering some uncertainty. Box-whisker plot technique has been used to detect outliers and summarize data to show the centrality, spread and skewness of data along Tigris river monitoring stations. The results showed that about 5% of data classified as outliers. An analysis of data by using percent of error in cation-anion balance technique to check reliability of data showed that 21.6 % of data exceeded the permit level of error in ions balance. Relative TDS/EC ratio accuracy check shows that the data agree with the range 0.55 to 0.75 with exception in stations from T11 to T16B where the ratio slightly less than 0.55. Also the results showed that they are not entirely consistent with the nature and the characteristics of the river water quality especially at the first section about 400 km from T1 to T10 where the bicarbonate is the domain anion and final section about 300 km from T29 to T33 where sulphate is the domain anion. Analysis of

Data showed that the data results have serious measuring or sampling errors, which means that the resultant data quality is insufficient for drawing reliable conclusions about water quality and for supporting decision making with high level of confidence. Applying quality control and quality assurance procedures have been required to ensure validity and reliability of data.

Keyword: box-whisker plot, cation-anion balance, data reliability, TDS/EC Ratio, water quality,

1- Introduction

Data are a vital component of water resources management activities performed to support water quality. Data collected by monitoring program are used to assess water conditions, investigate specific water quality issue or determine long-term trends. Monitoring program has been used to support decision making to sustain water quality. To make sure that all data generated by monitoring program are reliable, quality assurance and quality control should be included in program designs for all water quality monitoring activities. Data quality, however, fundamentally depends on the intended use of the data. To be meaningful, the data quality must meet the desired level of confidence for the purpose of the sampling program. As well, the sampling design and data quality should be able to perform over a wide range of possible outcomes [1]. The consequences of using poor quality data include faulty decisions, higher risk to the environment or human health, wasted resources, loss of credibility and sometimes, legal liability[2].

Uncertainty is introduced into measured water quality data by sample collection, preservation and analysis; therefore these activities must follow recommended (standard) procedures. All data are wasted if the collection, conservation and analysis of water sample are not done correctly and precisely [3]. In addition, an appropriate quality assurance protocols must be adhered to, in order to have confidence in the results and reduce the errors. Undetermined errors cause variations above or below the actual values; they are inconsistent and often small and thus are difficult to detect [4]. Evaluation data validity and reliability are important because measured data uncertainty effects water quality assessment [5].

Ministry of Environment has sampled water quality in rivers, lakes, streams and other resources. The data of parameters measurements were stored in unorganized excel spreadsheet. These data are often the result of a variety of calculations, manipulations, and transformations that are not evident in the numbers themselves. This can lead to several problems. The important issue here is that these data should be accurate and reliable to lead to sound decisions on management of water resources quality [6]. However, data users and decision-makers still suffer from poor information when they attempt to use the available data. One of the major reasons for this situation is that the basic requirements needed to ensure the accuracy and reliability of data is often overlooked. In addition, data validation in particular is poorly achieved. This work is an attempt to evaluate the validity and reliability of water quality data which collected by water resources monitoring program to ensure that the data can be used for decision making in the management of water resources with high level of confidence, ensure the validity of water quality data and improve the reliability of water quality assessments, and discovering some uncertainty.

2- Materials and Methods

The data used in this paper were obtained from water quality control department- Iraqi Ministry of Environment for the period from January 2013 to December 2013 which represents the monthly average values for thirteen water parameters covered by monitoring program. These parameters include (pH, DO, Ca, Mg, Na, K, SO_4, Cl, Alk., NO_3, PO_4, TDS and EC). These parameters were measured at thirty nine predefined points (stations) as shown in Table 1 along Tigris River to appropriately reflect the water condition within the river. Data results from about 468 water samples collected throughout the above period are subjected to different techniques to check the reliability and validity of collected water quality data.

Box-whisker plots are an ideal tool for evaluating differences between two or more groups of environmental data (generally for the same parameter) and obtaining outliers' data. They are also useful for examining data spread, central tendency, skewness, and the presence or absence of outliers [7]. The type of box plot used in this analysis is the standard box plot. The box itself contains the center 50 percent of the data (i.e., the interquartile range), and the median is indicated as a horizontal line within the box. The top edge of the box is the 75th percentile and the bottom edge is the 25th percentile. Vertical lines, sometimes called whiskers, extend to the last observation within one step beyond either end of the box. A step is 1.5 times the height of the box. Data points that fall outside one step are considered to be "outliers" [8].

The cations-anions balance technique is conducted to test validation and reliability of water quality data for the major ionic species. The principle of electroneutrality requires that the sum of the positive ions (cations) must equal the sum of the negative ions (anions). The error in cation-anion balance can be calculated by equation below [9, 10] where the ions are expressed in meq/l:

$$\text{Ion balance error \%} = \left(\frac{\sum cations - \sum anions}{\sum cations + \sum anions}\right) \times 100$$

According to Hounslow (1995) [3], the electrical balance error should be less than ±5% for good measurement and if it is greater than ± 5%, the analysis is supposed to be poor. But for surface water up to ± 10% is acceptable [11, 12]. If it is greater, the analysis does not pass the validation check.

Another approximate accuracy check is possible using the EC and TDS determinations. The TDS (in mg /l) should be between 0.55 and 0.75 times the EC (in μS/cm) for most waters up to a TDS of few thousand mg/ l [3].

3- Results and Discussions

According to American Public Health Association (APHA, 2005) [13], theoretically no result should be rejected, because it may indicate either an analytical problem that casts doubt on all results, or the presence of a true variant in the distribution of the data. In practice, reject the result of any analysis in which a known error has occurred. In environmental studies, extremely high and low concentrations of contaminants may indicate the existence of areas with problems or areas with no contamination, so they should not be rejected arbitrarily. Worthwhile it is ascertained that the data must gained from a series of systematic or standard procedures with quality control and quality assurance procedures to ensure reliability and validity of data.

3.1- Box-whisker plot technique and outliers

Statistical analysis and techniques for data such as technique of Box-Whisker plots are used to provide summarization and detect outliers in data sets. Data of thirteen water quality indicators are analyzed by Box-whisker plots. As shown in Figures (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and 13) box-plots are used to summarize water quality indicators data and detect outlier in data sets of pH, DO, Ca, Mg, Na, K, SO_4, Cl, Alk., NO_3, PO_4, TDS and EC respectively. As shown in these Figures there are differences between the spread of data of groups (stations) along Tigris River. In general, data spread shows increasing especially at sampling station number T16B and downstream the river with some exception. Increasing data spread means high variability (high standard deviation) of data which indicate low precision of data (analytical results) [14, 3]. About 5 % (300 out of 6360 measurements) of data analyzed are classified as outliers' data. Extreme or outlying values can lead to inaccurate results [15]. An outlier may indicate bad data. On the other hand, in some cases, outliers may be due to random variation or may indicate something scientifically interesting [8]. Data distribution as shown in Figures (1 to 13) exhibits skewed distribution. Using median instead of mean is more reliable to show the centrality of data, because it is more resistant than mean in presence of outliers and skewed distribution [16].

3.2- Cation-Anion Balance Technique

When all the major cations (such as Ca, Mg, Na, K) and anions (such as Cl, SO_4, HCO_3 and NO_3) have been analyzed, one of the most important validation tests can be conducted is the cation-anion balance technique. Applying this valid check technique as shown in Table (2) as sample of calculations, on data of 468 water samples collected from 39 stations along Tigris River in 2013, exhibits that 21.6 % of data (83 out of 468 measurements) exceeded the acceptance limit (± 10 %) of cation-anion balance error percent. This indicates that around one fifth of the samples have serious measuring or sampling errors, which means that the resultant data quality is insufficient for drawing reliable conclusions about water quality. Deviations from electroneutrality come from random or systematic error in one or more constituents or an incomplete analysis that neglects some significant constituent [10].

3.3- comparison of EC and TDS

The relationship of total dissolved solids to electrical conductivity is a useful accuracy check for reliability of measurements (data). For water of ordinary composition, the dissolved-solids value in milligrams per liter should generally lie between 0.55 to 0.75 times the specific conductivity in microsemens per centimeter. Generally, analysis of available data has shown that there is variability. As shown in figure 14, TDS relative to EC does not agree with this range at stations from T11to T16B, where the ratio is slightly less than o.55. Data of the other stations showed agreement with this range. Water in which anions are mostly dominated by bicarbonate and chloride should have a factor near the lower end of this range whereas waters high in sulphate may reach or even exceed the upper end [10]. The results are not entirely consistent with the nature and the characteristics of the river water quality as shown in figure 14 especially at the first stations from T1 to T10, where the bicarbonate is the domain anion and final stations from T29 to T33, where sulphate is the domain anion. For repeated samples from the same source, a well-defined relationship of dissolved solids to conductivity often can be established, and this can afford a good general accuracy check for analyses of these samples.

4- Conclusions

Data analysis results showed that there is about 5 % of data analyzed are classified as outliers' data. Applying ions balance validity check exhibits that 21.6 % of data exceeded the acceptance limit (± 10 %) of cation-anion balance error percent. This indicates that the data results have serious measuring or sampling errors, which means that the resultant data quality is insufficient for drawing reliable conclusions about water quality and for supporting decision making with high level of confidence. Applying quality control and quality assurance procedures have been required to ensure validity and reliability of data.

Table (1): Number of monitoring stations at each Province along Tigris River within Iraq.

Number of Stations	Stations code	The Province
10	T1, T2, T3, T4, T5, T6, T7, T8, T9 and T10	Nineveh
7	T11, T12, T12S, T13, T14, T15 and T16	Salah Al- Dien
11	T16B, T17, T18, T19, T20, T21, T22, T23, T24, T24B1 and T24B2	Baghdad
4	T25, T26, T27 and T28	Kut
6	T29, T29M, T30, T31, T32 and T33	Maysan
1	T34	Basra

Table (2): Sample of calculation of ion balance error % at station No. T2

Cations	mg/l	meq/l	Anions	mg/l	meq/l
Ca	58.9	2.939	Cl	19	0.536
Mg	17.3	1.418	SO4	46.5	0.968
Na	11.8	0.513	HCO3	135	2.212
K	1.75	0.045	NO3	0.85	0.014
∑ cations = 4.915			∑anions = 3.730		
Ion balance error % = $\left(\frac{\sum cations - \sum anions}{\sum cations + \sum anions}\right) \times 100$ = **13.71** %					

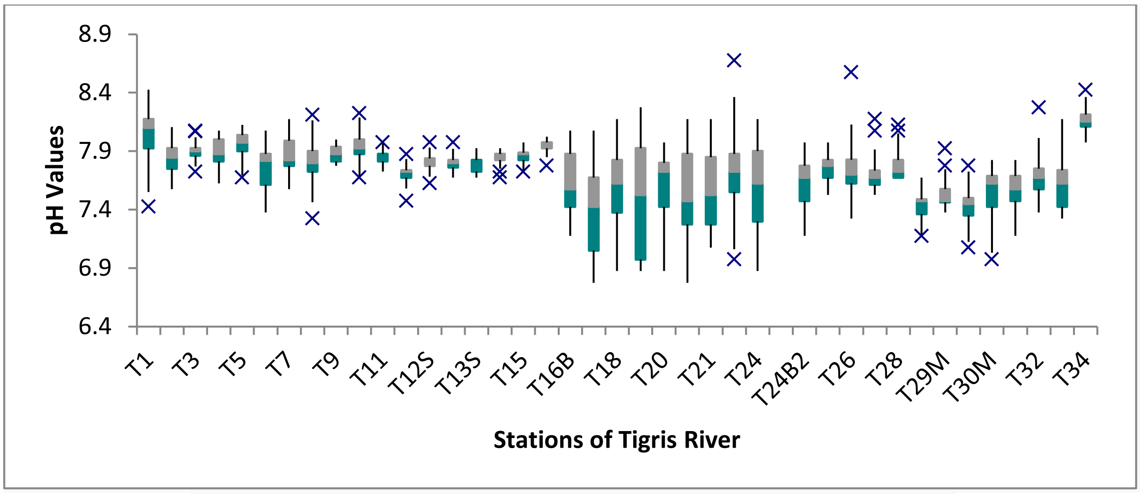

Figure1: Box-plot of pH variation along stations of Tigris River.

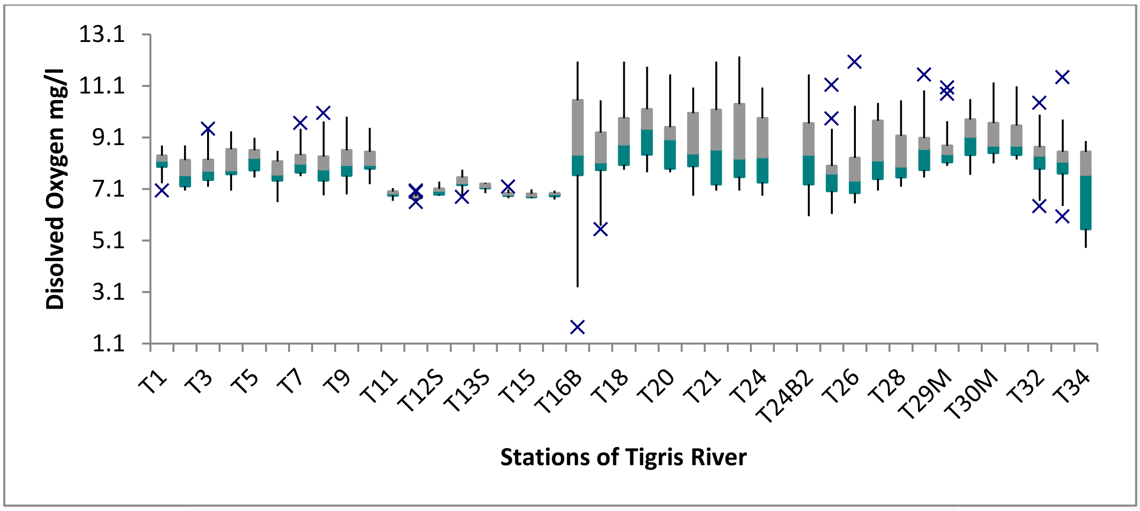

Figure 2: Box-plot of DO variation along station of Tigris River.

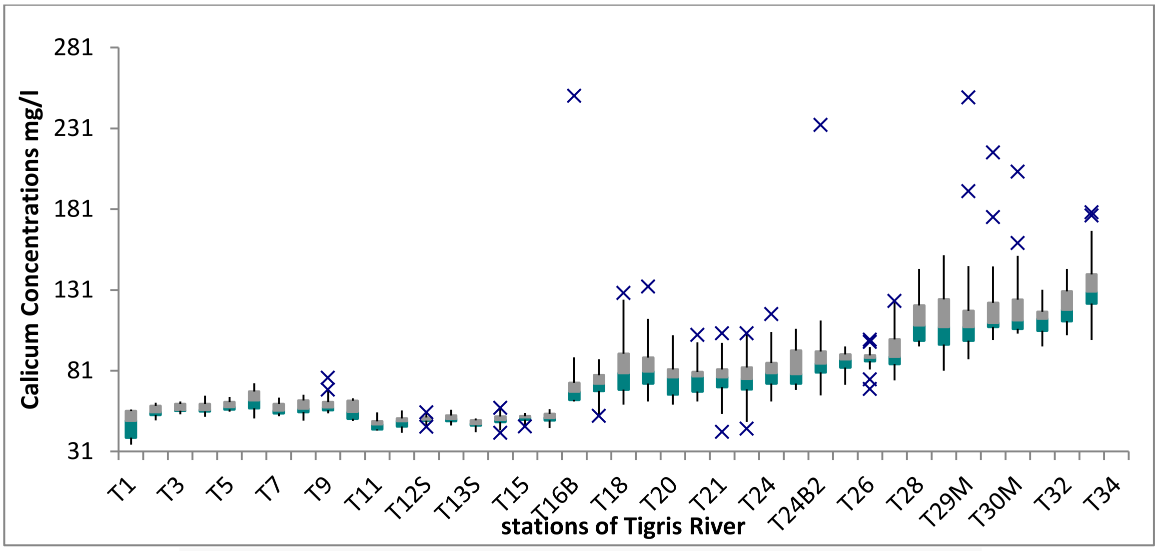

Figure 3: Box-plot of Ca ion variation along station of Tigris River.

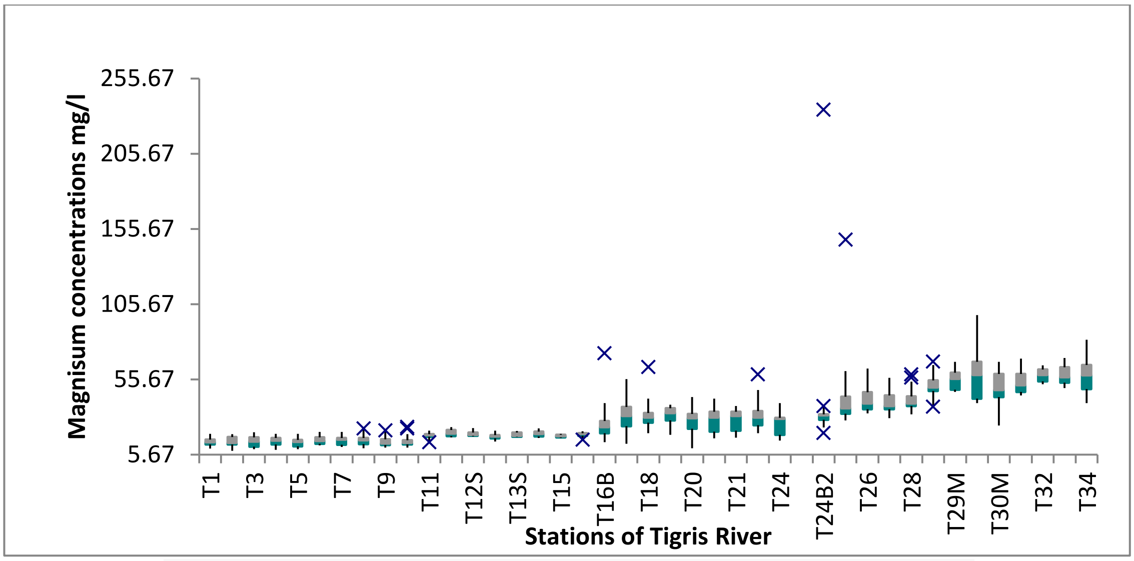

Figure 4: Box-plot of Mg ion variation along station of Tigris River.

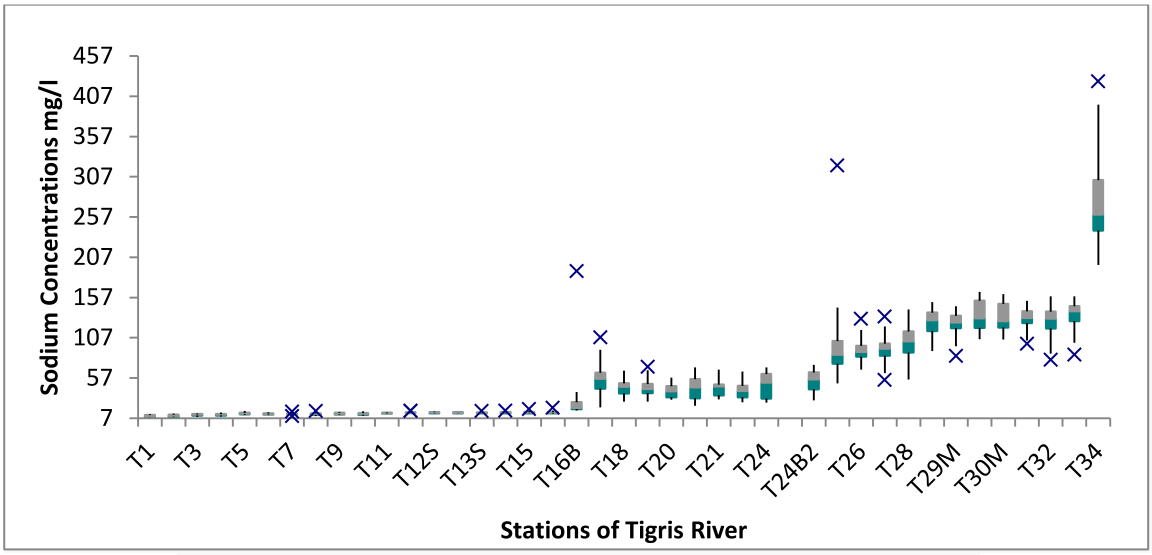

Figure 5: Box-plot of Na ion variation along station of Tigris River.

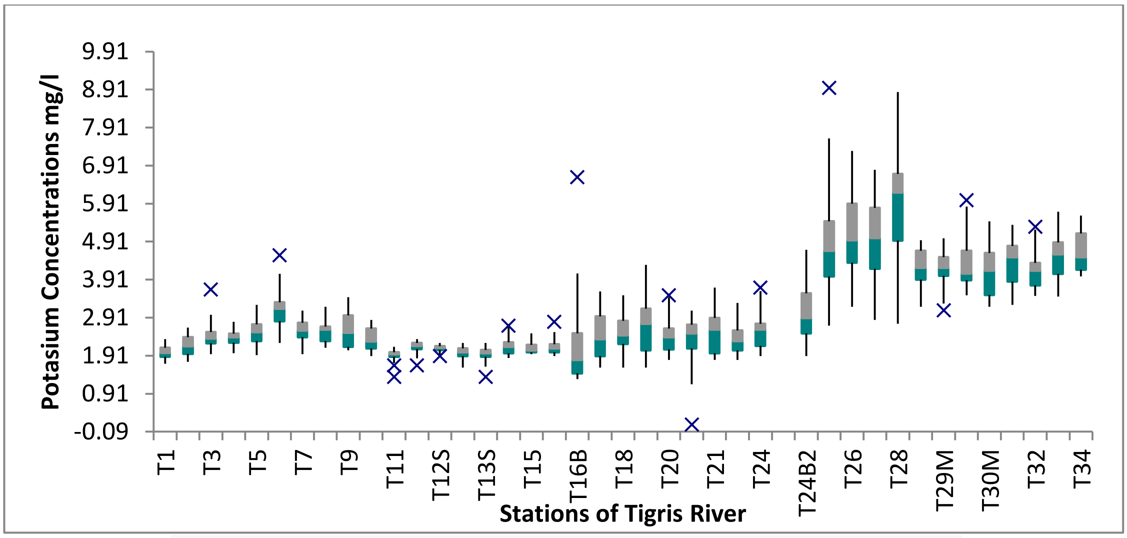

Figure 6: Box-plot of K ion variation along station of Tigris River.

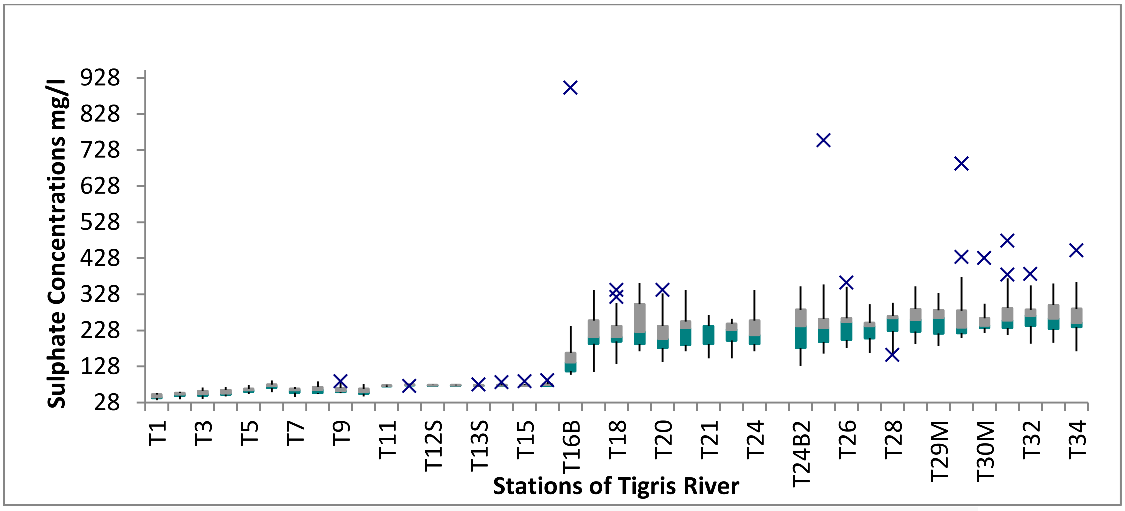

Figure 7: Box-plot of SO_4 ion variation along stations of Tigris River.

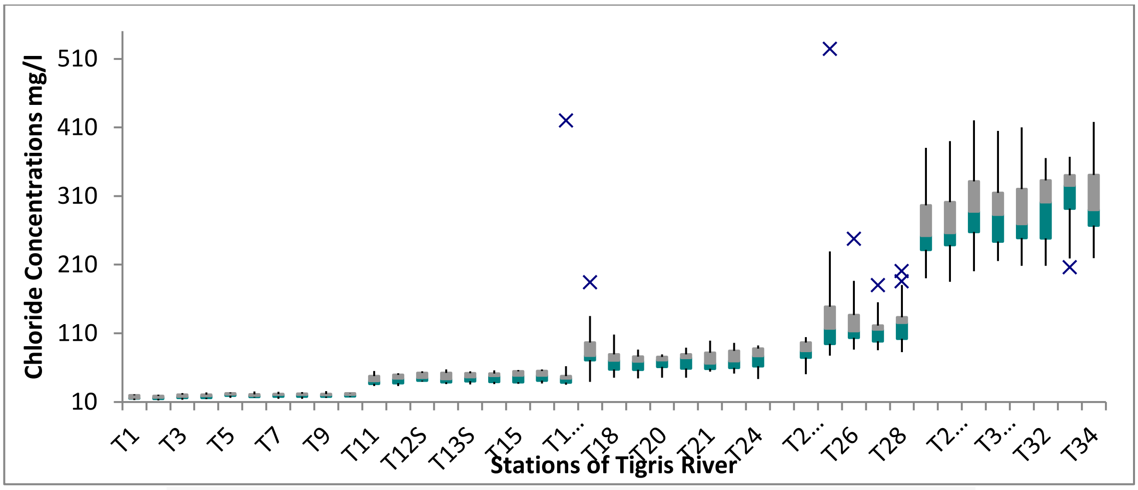

Figure 8: Box-plot of Cl ion variation along stations of Tigris River.

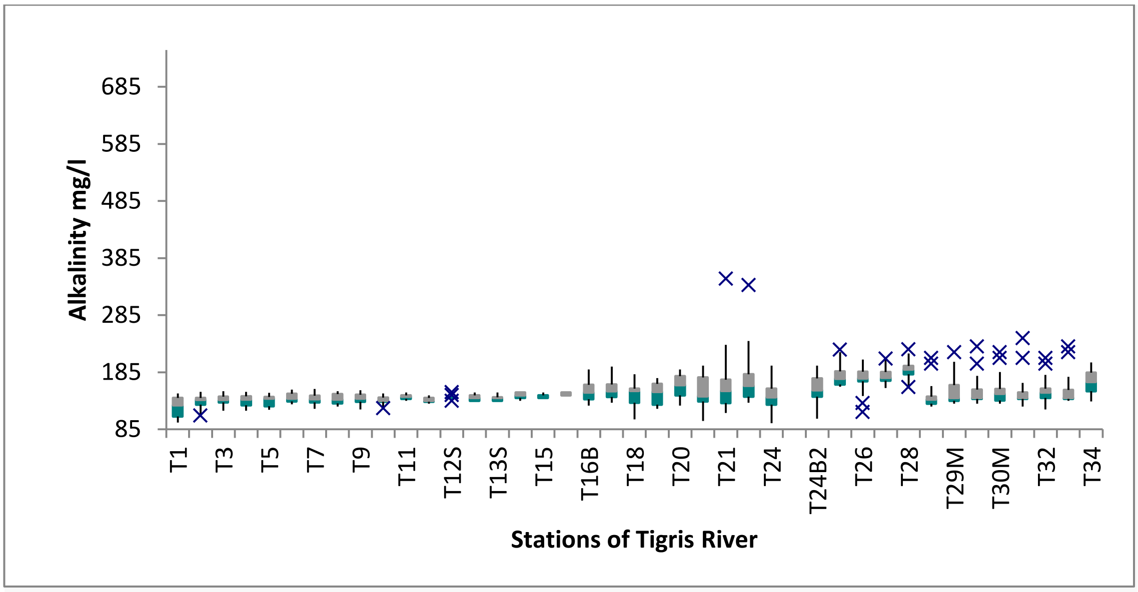

Figure 9: Box-plot of Alkalinity variation along stations of Tigris River.

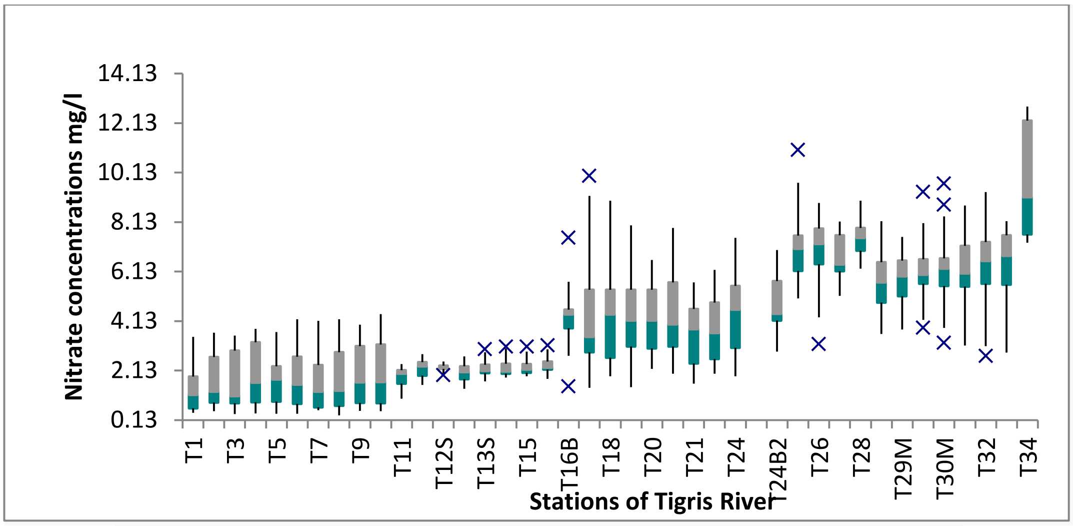

Figure 10: Box-plot of NO_3 ion variation along stations of Tigris River.

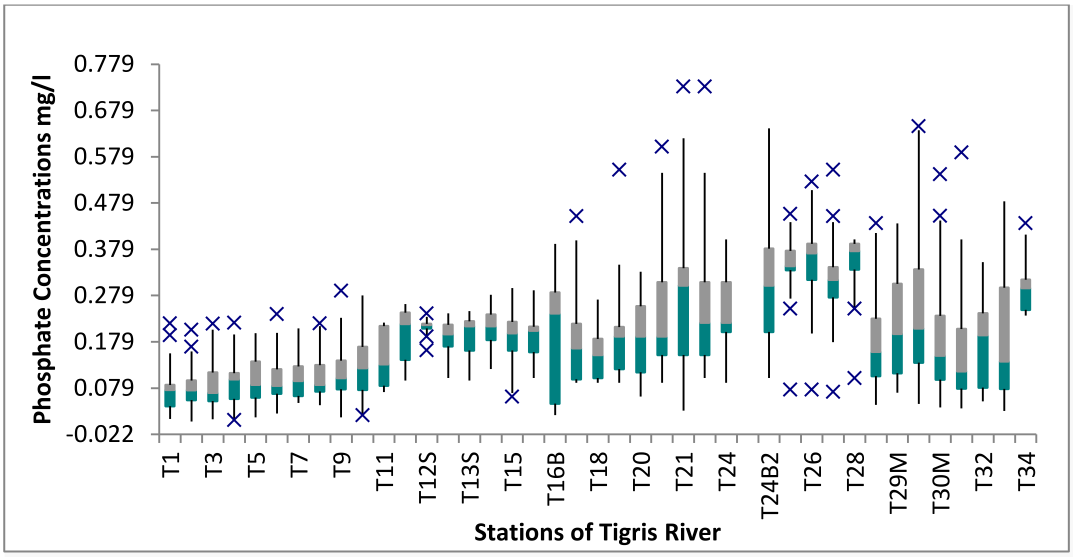

Figure 11: Box-plot of PO_4 ion variation along stations of Tigris River.

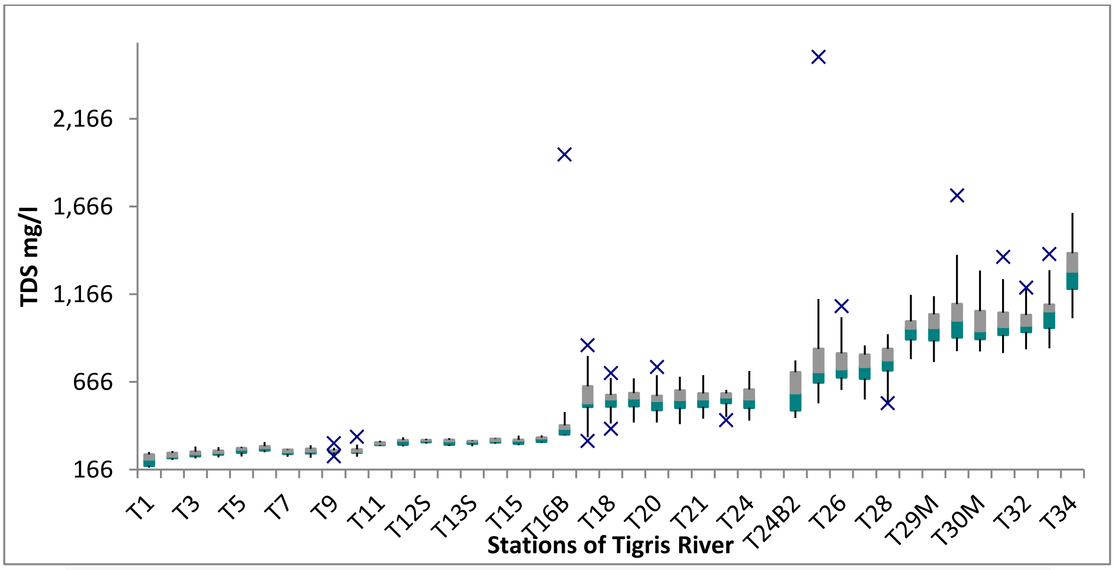

Figure 12: Box-plot of total dissolved solid variation along stations of Tigris River.

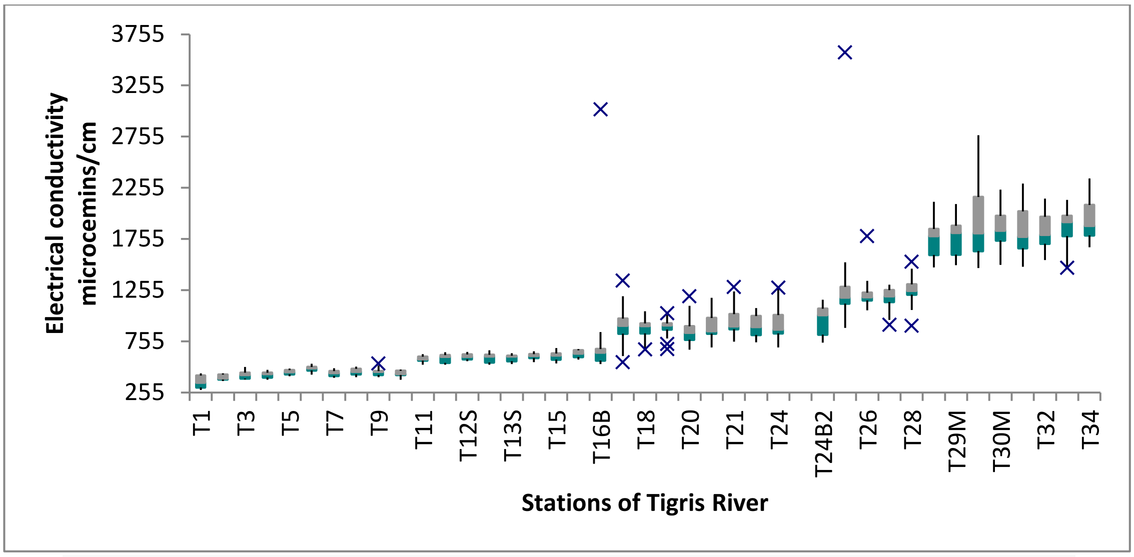

Figure 13: Box-plot of electrical conductivity variation along stations of Tigris River.

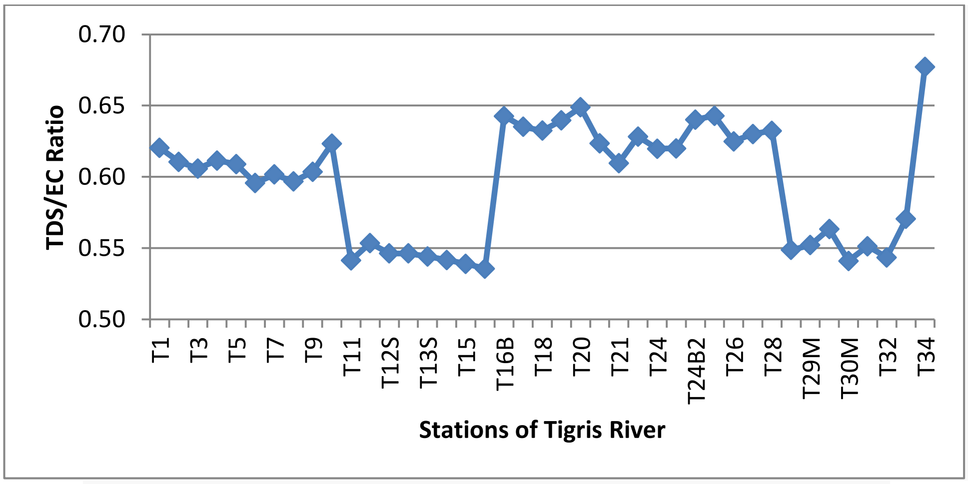

Figure 14: Variation of TDS/EC ratio along stations of Tigris River.

References

[1] Mitchell, Patricia, 2006. Guidelines for Quality Assurance and Quality Control in Surface Water Quality Programs in Alberta, Alberta Environment,Pub. No.T/884, Canada.

[2] Lombard, S. and C. Kirchmer. 2004. Guidelines for preparing quality assurance project plans for environmental studies. Washington State Dept. of Ecology, Pub. No. 04-03-030, 47 p. + app.

[3] Hounslow, Arthur W., 1995. Water quality data: analysis and interpretation. Lewis Publisher, CRC Press LLC, USA, p.49, p.73.

[4] Carneiro, Charles, Andreoli, Cleverson V., Cunha, Cynara, L., Gobbi, Eduardo F., 2014. Reservoir Eutrophication: preventive management- an applied example of integrated basin, IWA Publishing, UK, p.164.

[5] Harmel, R. D., R. J. Cooper, R. M. Slade, R. L. Haney, and J. G. Arnold. 2006. Cumulative uncertainty in measured stream flow and water quality data for small watersheds. Trans. ASABE 49(3): 689-701.

[6] Harmancioglu. N. B., 2007. Informed decision making for water resources management, Proceeding of DSI and WWC International Congress on River Basin Management, Antalya, Turkey, March 2007.

[7] USEPA. 1992. A Ground Water Information Tracking System with Statistical Analysis Capacity. United States Protection Agency, Region VII, Kansas City, Kansas.

[8] Liya Fu and You-GanWang (2012). Statistical Tools for Analyzing Water Quality Data, Water Quality Monitoring and Assessment, Dr. Voudouris (Ed.), ISBN: 978-953-51-0486-5, InTech, Available from: http://www.intechopen.com/books/water-quality-monitoring-and- assessment/statistical-tools-for-analyzingwater- quality-data.

[9] Freeze, R. A. and J.A. Cherry, 1979. Groundwater, Prentice-Hall Inc., Englewood Cliffs, NJ, p.97.
[10] Hem, J D. 1985. Study and interpretation of the chemical characteristics of natural water. United States Geological Survey Water-Supply Paper 2254.p.165
[11] Katz, Brian G. and Collins, Jerilyn J.,1998. Evaluation of Chemical Data from Selected Sites in the Surface-Water Ambient Monitoring Program (SWAMP) in Florida, Open-File Report 98-559, USGS, p.4.
[12] Abreha, Asmelash G., 2014. Hydrogeochemical and Water Quality Investigation on Irrigation and Drinking Water Supplies in the Mekelle Region, Northern Ethiopia, Msc. Thesis, Faculty of Geo-Information Science and Earth Observation ,University of Twente, p.24.
[13] APHA, 2005, American Public Health Association, American Water Works Association, Water Environment Federation. Standard methods for the examination of water and wastewater, 21st Ed. Washington DC, USA.
[14] John R. Taylor, 1997. An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, 2d Edition, University Science Books, USA.
[15] József Kovács, Péter Tanos, János Korponai, Ilona Kovácsné Székely, Károly Gondár, Katalin Gondár-Sőregi and István Gábor Hatvani (2012). Analysis of Water Quality Data for Scientists, Water Quality Monitoring and Assessment, Dr. Voudouris (Ed.), ISBN: 978-953-51-0486-5, InTech, Available from: http://www.intechopen.com/books/water-quality-monitoring-and-assessment/analysis-of-water-quality-data-forresearchers
[16] Helsel, D. R. and Hirsch, R. M., 2002. Statistical Methods in Water Resources-Techniques of Water-Resources Investigations of the United States Geological Survey, book4-Ch a3, USGS.

موثوقية بيانات برنامج رصد جودة الموارد المائية

صباح عبيد حمد الشجيري
دائرة حماية وتحسين البيئة في المنطقة الوسطى- وزارة البيئة- العراق
البريد الالكتروني: dr.sabah2012@gmail.com .

الخلاصة:
تعد البيانات ذات اهمية قصوى في فعاليات ادارة الموارد المائية. ان تداعيات استخدام بيانات ضعيفة وغير موثوقة تشمل القرارات الخاطئة، ارتفاع المخاطر على البيئة وصحة الانسان، اهدار الموارد وفقدان المصداقية. تقنية الرسم الصندوقي وتقنية حساب نسبة الخطأ في توازن الأيونات ونسبة الأملاح الذائبة الى الموصلية الكهربائية استخدمت لاختبار صحة بيانات جودة المياه. هذه الطرق طبقت على بيانات جودة المياه لنهر دجلة لعام 2013 للتأكد من أن البيانات يمكن أن تستخدم لصنع القرار في إدارة جودة الموارد المائية مع مستوى عال من الثقة، وتحسين موثوقية تقييم نوعية المياه، واكتشاف بعض الأشكالات. استخدمت تقنية الرسم الصندوقي للكشف عن القيم الشاذة وتلخيص البيانات لاظهار التمركز والانتشار والالتواء للبيانات لمحطات الرصد على طول مجرى النهر. أظهرت النتائج بهذه الطريقة أن حوالي 5 % من البيانات صنفت على انها شاذة. تحليل البيانات باستخدام تقنية نسبة الخطأ في توازن الايونات الموجبة والسالبة للتاكد من موثوقية البيانات قد اظهر ان 21.6 % من البيانات المفحوصة تجاوزت النسبة المئوية المسموحة للخطا في توازن الايونات. أن نسبة الاملاح الذائبة الكلية الى الموصلية الكهربائية كطريقة تقريبية للتأكد من دقة البيانات اظهرت ان النتائج تتوافق مع النسبة التي تتراوح بين (0.55- 0.75) الا بعض الاستثناء في المحطات (T11- T16B) حيث تنخفض النسبة قليلا عن 0.55 . ان النتائج رغم انها ضمن النسبة ولكنها لاتتسق تماما مع طبيعة وخصائص نوعية المياه في مقاطع النهر, حيث أن المياه في اعلى النهر غنية بايونات البيكربونات ويفترض ان تكون النسبة بالحد الادنى، ونوعية المياه في اسفل النهر غنية بايونات الكبريتات ويفترض ان تكون النسبة بحدها الاعلى وهو مالم يتوافق تماما مع النتائج. خلاصة القول ان نتائج تحليل البيانات اظهر آنها تحتوي على أخطاء جسيمة ربما تتعلق بالنمذجة اوالقياسات وهو مايعني ان نوعية البيانات الناتجة بحاجة الى مراجعة وغير كافية لاستخلاص استنتاجات موثوقة حول تقييم نوعية المياه ودعم اتخاذ القرار على مستوى عال من الثقة. ان ضمان صحة وموثوقية البيانات يتطلب تطبيق برنامج او اجراءات مراقبة الجودة وضمان الجودة بشكل فاعل وفي جميع المراحل.

الكلمات المفتاحية: الرسم الصندوقي، توازن الايونات الموجبة والسالبة، موثوقية البيانات، نسبة الاملاح الذائبة الى الموصلية الكهربائية، جودة المياه.