

Electrical Conductivity as a General Predictor of Multiple Parameters in Tigris River Based on Statistical Regression Model

Omar Anmar Almoula *
M.Sc. student
College of Engineering University of Baghdad
o.anmar1101@coeng.uobaghdad.edu.iq

Dr. Khalid Adel Abdulrazzaq
Assistant Professor
College of Engineering University of Baghdad
aleoubaidy@coeng.uobaghdad.edu.iq

Dr. Athraa Hashim Mohammed
Assistant Professor
College of Engineering University of Baghdad
athram2004@yahoo.com

ABSTRACT

Surface water samples from different locations within Tigris River's boundaries in Baghdad city have been analyzed for drinking purposes. Correlation coefficients among different parameters were determined. An attempt has been made to develop linear regression equations to predict the concentration of water quality constituents having significant correlation coefficients with electrical conductivity (EC). This study aims to find five regression models produced and validated using electrical conductivity as a predictor to predict total hardness (TH), calcium (Ca), chloride (Cl), sulfate (SO₄), and total dissolved solids (TDS). The five models showed good/excellent prediction ability of the parameters mentioned above, which is a very good startup to establish a rule of thumb in the laboratories to compare between observations. The importance of linear regression equations in predicting surface water quality parameters is a method that can be applied to any other location.

Keywords: Tigris River, Electrical conductivity, Regression.

التوصيلية الكهربائية كمتوقع عام لمعلومات متعددة في نهر دجلة على اساس نموذج الانحدار الاحصائي

د. عذراء هاشم محمد
استاذ مساعد في كلية الهندسة / جامعة بغداد

د. خالد عادل عبدالرزاق
استاذ مساعد في كلية الهندسة / جامعة بغداد

*عمر انمار خليف المولى
طالب دراسات عليا / ماجستير

الخلاصة

تم تحليل عينات المياه السطحية من مواقع مختلفة ضمن حدود نهر دجلة في مدينة بغداد لأغراض الشرب. تم تحديد معاملات الارتباط بين المتغيرات المختلفة. جرت محاولة لتطوير معادلات الانحدار الخطي للتنبؤ بتركيز مكونات جودة المياه التي لها معاملات ارتباط كبيرة مع التوصيلية الكهربائية. ان الهدف من هذه الدراسة هو الحصول على خمسة نماذج لتوقع العلاقات والتحقق منها باستخدام التوصيلية الكهربائية كمتوقع للتنبؤ بكل من العسرة الكلية، الكالسيوم، الكلوريدات، الكبريتات و المواد

*Corresponding author

Peer review under the responsibility of University of Baghdad.

<https://doi.org/10.31026/j.eng.2021.02.06>

2520-3339 © 2019 University of Baghdad. Production and hosting by Journal of Engineering.

This is an open access article under the CC BY4 license <http://creativecommons.org/licenses/by/4.0/>.

Article received: 18/7/2020

Article accepted: 20/9/2020

Article published:1/2/2021



الصلابة الذاتية الكلية. أظهرت النماذج الخمسة قدرة تنبؤية جيدة / ممتازة للمتغيرات المذكورة أعلاه وهي بداية جيدة جدًا لتأسيس قاعدة أساسية في المختبرات للمقارنة بين المتغيرات. أهمية معادلات الانحدار الخطي في التنبؤ بمعلمات جودة المياه السطحية كطريقة يمكن تطبيقها في مواقع أخرى.
الكلمات الرئيسية: نهر دجلة، التوصيلية الكهربائية، التوقع.

1. INTRODUCTION

It is shown that there is strong evidence that electrical conductivity is the most appropriate variable predicting or explaining more than values of the dependent variables (**Joarder et al., 2008**). Sustainable development is closely linked to obtaining adequate quantitative and qualitative water to maintain environmental and health systems. It is essential for human, economic, and social development despite the increasing water demand is spotted through expanded international demand and mismanagement of water resources (**Bhaduri et al., 2016**).

In Iraq, river water is being used for different purposes (domestic, agriculture, and industrial); water quality could be a function of its physicochemical properties. Continuous monitoring of a large number of quality parameters is crucial for the efficient maintenance of water quality. Nevertheless, it is a complex and challenging responsibility for routine monitoring of all the parameters, even if satisfactory personnel and laboratory resources are available. Hence, another approach based on correlation and Regression has recently been deployed to improve empirical relationships for comparison between physicochemical parameters (**Bhandari and Nayal, 2008**).

Electrical Conductivity (EC) and total dissolved solids (TDS) are correlated and applied to portray the extent of salinity. The water TDS experiment is more complicated than that of EC; nonetheless, the TDS test is imperative since it is superior to EC in illustrating groundwater quality, specifically in the influence of seawater intrusion. These circumstances create the need for research in TDS/EC correlation and Regression (**Rusydi, 2018**).

(**Bhandari and Nayal, 2008**), studied a broad spectrum of water quality-related correlations among them, EC was correlated to total hardness and sulfate, and total dissolved solids with R equals 0.89, 0.806, and 0.94 respectively. (**Kumar and Sinha, 2010**), showed that EC is having a noteworthy correlation with several quality parameters, including (total hardness, calcium, total dissolved solids, total solids, and sulfate) with a correlation coefficient (r) over 0.9.

In the present study, statistical regression modeling is adopted to produce five effective predictive models in which electrical conductivity is used as an independent variable (predictor) while the other five criterion variables (dependent variables) include: total hardness, calcium, chloride, sulfate, and total dissolved solids.

2. MATERIAL AND METHODS

2.1 Case study

The Tigris River in Baghdad city's boundaries has been considered and represents the only source of drinking water. The length of the river is about 65 km, and it divides the capital into two parts of Al-Karkh (west bank of a river) and Al-Rusafa (east bank of a river), nine monitoring stations initially constructed near the water treatment plant (WTP) intakes (**Khudair, 2013**). Each station has its own laboratory, and the observations taken from these stations are considered independent. The case study is shown in **Fig. 1**; where Al-Karkh Station (S1), East Tigris Station (S2), Al-Sader Station (S3), Al-Karama Station (S4), Al-Wathba Station (S5), Al-Qadisiyah Station (S6), Al-Dora Station (S7), Al-Wahda Station (S8), and Al-Rashed Station (S9).

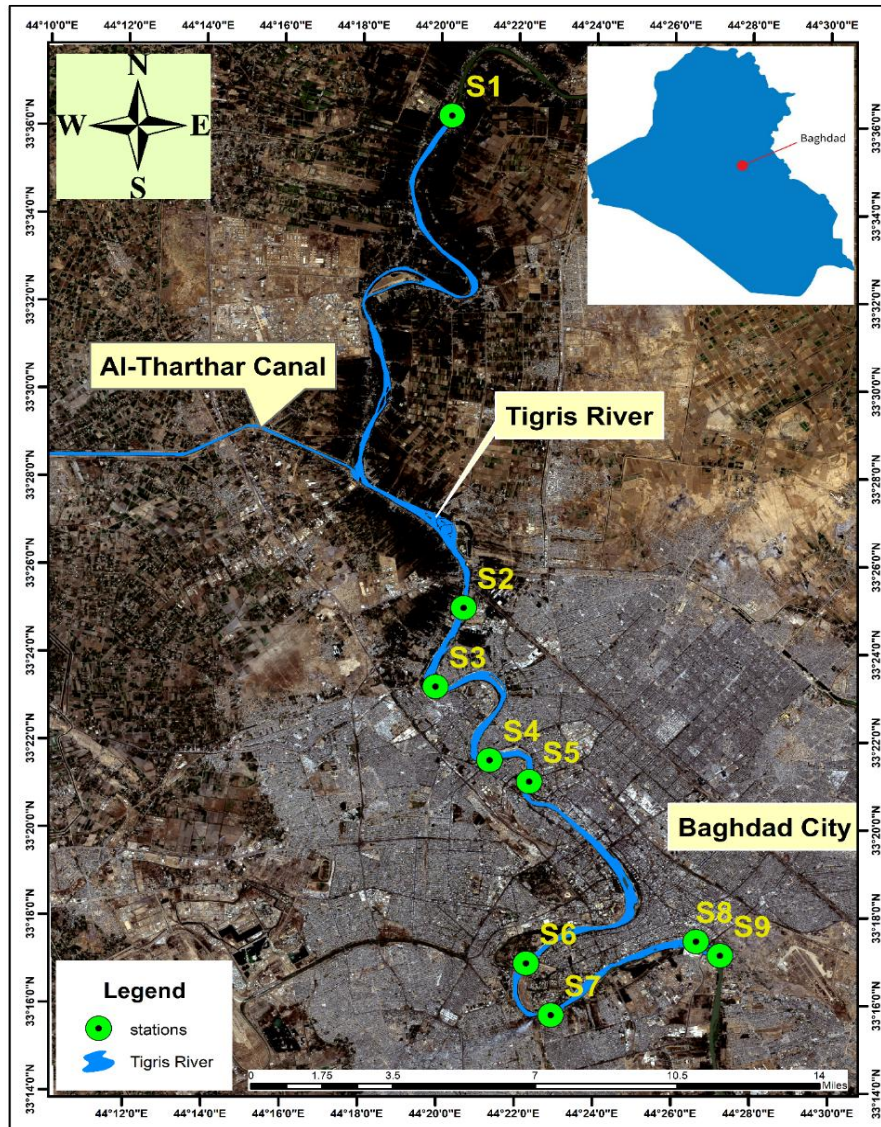


Figure 1. The study area map for the stations along the Tigris River in Baghdad city.

2.2 Data collection and analysis

In the early stage of statistical analysis, 15 different parameters were considered. The data are collected from Baghdad mayoralty in coordination with the different WTPs; the study period is selected for the period (2014- 2019). After reviewing the data and cross-correlation between the parameters; Electrical conductivity showed a good correlation with Total Hardness (TH), Calcium (Ca), Chloride (Cl), Sulfate (SO₄), and Total Dissolved Solids (TDS). Hence, the dataset is prepared, and the descriptive statistics in **Table 1**. are reported to understand the nature of the data. It is noteworthy that all test procedures are done according to standard procedures (APHA, 2017). It is noteworthy in **Table 1**. that the mean value of TDS exceeded the WHO standards while all other parameters are within the acceptable limits according to the world health organization.



Table 1. Descriptive statistics of Tigris River water quality parameters.

| Parameter | N | Minimum | Maximum | Mean | WHO* | STDV |
|--------------------------------|-----|---------|---------|------|------|--------|
| EC (µS/cm) | 610 | 528 | 1222 | 881 | 1000 | 150.33 |
| TDS (mg/L) | 610 | 323 | 838 | 588 | 500 | 104.85 |
| TH (mg/L) as CaCO ₃ | 610 | 223 | 456 | 328 | 500 | 53.56 |
| Ca (mg/L) | 609 | 52 | 129 | 82.5 | 200 | 16.97 |
| Cl (mg/L) | 610 | 24 | 121 | 70.9 | 250 | 16.92 |
| SO ₄ (mg/L) | 607 | 56 | 363 | 206 | 250 | 59.8 |

* World health organization standards (WHO, 2011).

2.3 Methodology

Correlation coefficients are a common method for characterizing the association between two variables. Correlation is confirmed when both variables have random measurement errors. The variables are equally important and interchangeable (both variables are correlated, but it does not mean that they are governed by predictor / predicted relation). Pearson sample correlation coefficient between x (first variable) and y (second variable) is given by (Al Jawad and Tariq, 2019; Berthouex and Brown, 2002).

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{1}$$

In this study, a simple linear regression model was adopted. Although data fit is a useful criterion for evaluating a model, it should not be concerned with obtaining the optimal mathematical fit (i.e., the adjusted maximum R² or the minimum residual standard deviation) of the estimated data. The simple linear regression model can be expressed in the general form (Kutner et al., 2005; Krishnan, 2006):

$$Y = C_0 + C_1 X \tag{2}$$

Where Y and X represent the dependent and independent variables, while C₀ and C₁ represent the y-intercept value and the slope of the line, respectively.

(Snee, 1977) had found the following procedures are useful in validating the regression model:

- Comparison of model (Y) predictions and coefficients (C) with physical theory. Negative expectations for theoretically positive quantity or false-marked transactions are indications of an inappropriate or poorly evaluated model.
- Collect new data to verify model predictions.
- Reserving a portion of the available data for an independent measurement of the model's prediction accuracy.

The collection of new data is the preferred method for validating the model. In many cases, this is neither practical nor possible. In this case, a procedure that simulates the collection of new data is needed. A sensible way forward is to split the data in one hand into two groups. The first data set, called the estimate data set, is used to estimate model parameters. The remaining data points, called the verification data set, are used to measure the accuracy of the model's prediction.



The Data splitting approach is used here, using a random sampling technique and split the dataset. The validation of the data splitting method is done using criteria estimates such as root mean square error RMSE given by:

$$RMSE_{val} = \sqrt{\frac{1}{n_{val}} \sum_{i=1}^{n_{val}} (\hat{y}_i - y_i)^2} \tag{3}$$

Where *val* subscript refers to the validation dataset.

3. RESULTS AND DISCUSSION

To perform the linear Regression analysis for the five parameters (TDS, TH, Ca, Cl, and SO₄) vs. EC, the data set containing more than 600 observations is separated into two data sets. The calibration dataset (used to create the models) and Validation dataset (used to check the models' performance for future prediction); in this study, the validation dataset is randomly taken from the whole data (60 observations are set aside for validation). Statistical program (SPSS) is deliberated in this study to perform the regression analysis. The summary of the models is shown in **Table 2**.

Table 2. Summary of regression models.

| Model Name | r | R ² | SE _{est} | C ₀ | C ₁ | RMSE _{val} |
|---------------------|-------|----------------|-------------------|----------------|----------------|---------------------|
| TDS-EC | 0.979 | 0.958 | 21.45 | -13.08 | 0.684 | 18.57 |
| TH-EC | 0.939 | 0.882 | 18.47 | 33.30 | 0.336 | 17.97 |
| Ca-EC | 0.874 | 0.764 | 8.31 | -4.90 | 0.100 | 8.89 |
| Cl-EC | 0.937 | 0.878 | 5.87 | -21.42 | 0.105 | 4.91 |
| SO ₄ -EC | 0.834 | 0.695 | 34.26 | -99.36 | 0.343 | 31.8 |

All significance value (p-value) were below 0.001, which indicate that the predictor parameter (EC) were significant to predict the criterion parameters (TDS, TH, etc.).

The R-Square indicates an excellent model for the TDS-EC model, explaining nearly 96% of TDS variance. Also, all other models are significant and well describe the behavior of the predicted parameter. The standard error of estimation (SE_{est}) is the average distance at which the observed values are from the regression line. Conveniently, it reports the average error of the regression model using dependent variable units. The smaller the SE_{est} value, the closer the values are to the regression line.

These models make physical sense as the positive sign of (C₁ and r) signify a direct relationship between the electrical conductivity and other parameters (for instance, when the EC increase, the Cl will increase). These results are analogous to (Joarder et al., 2008).

The physical explanation of the positive sign is that the EC in water increases as the positive and negative ion concentration increases (Peterson et al., 2013).

From **Table 2**., the equation form of the models can be stated:

$$TDS = -13.08 + 0.684 * EC \tag{4}$$



$$TH = 33.3 + 0.336 * EC \tag{5}$$

$$Ca = -4.9 + 0.1 * EC \tag{6}$$

$$Cl = -21.42 + 0.105 * EC \tag{7}$$

$$SO4 = -99.36 + 0.343 * EC \tag{8}$$

The predicted vs. observed plot is a powerful tool to emphasize the model's prediction ability, especially when the calibration and validation datasets are differentiated. **Fig.2, 3, 4, 5, and 6** show the prediction ability of the five models described above. For the TDS-EC model, it is evident that the model is very accurate as the points are concentrated along the diagonal line, while the SO₄-EC model is less accurate as the points are scattered along the diagonal line. Overall, the five models are acceptable and of high importance, as they indicate different parameters by observing the electrical conductivity only. Since the points are scattered, it indicates that the data meet the assumptions of the normally distributed errors and that the residue variations are constant. RMSEval has shown an acceptable value for all models as it is much lower than the STDV of the variable.

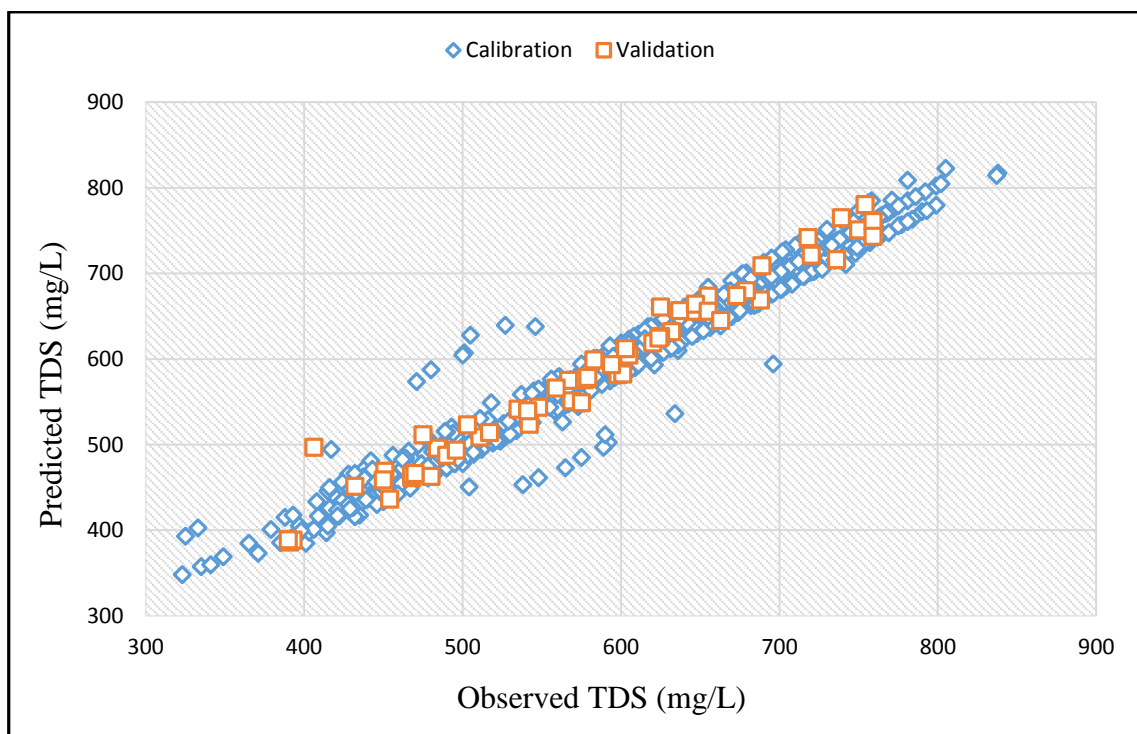


Figure 2. TDS-EC regression model.

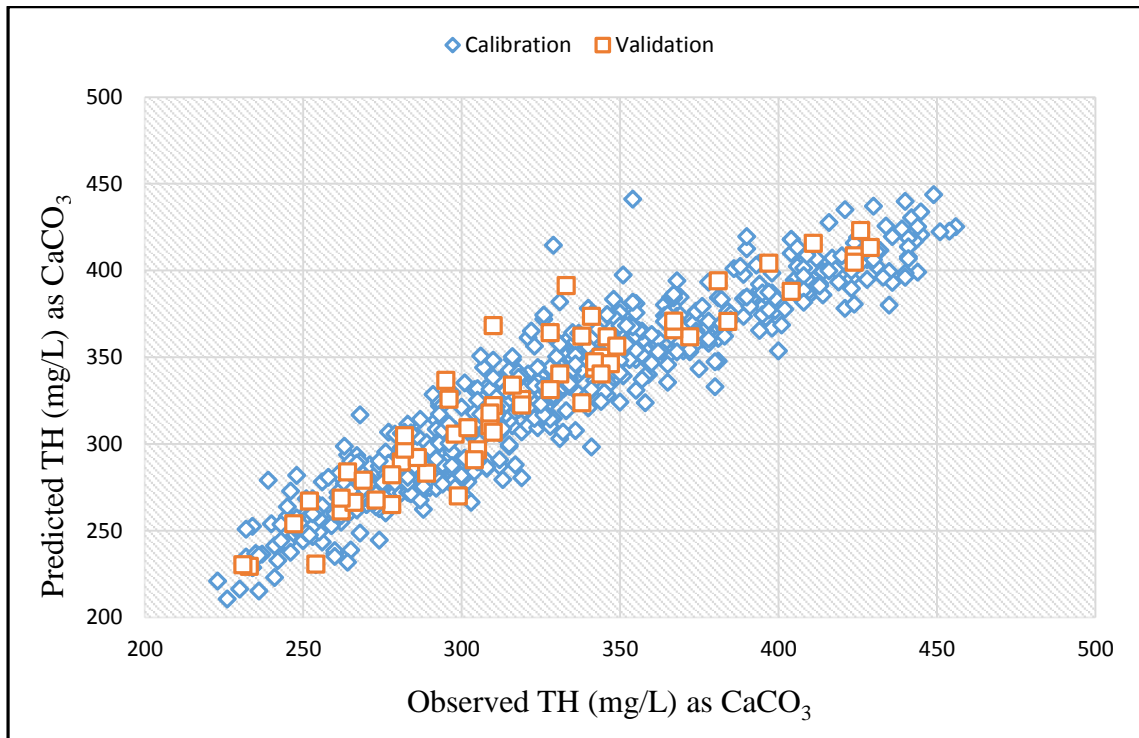


Figure 3. TH-EC regression model.

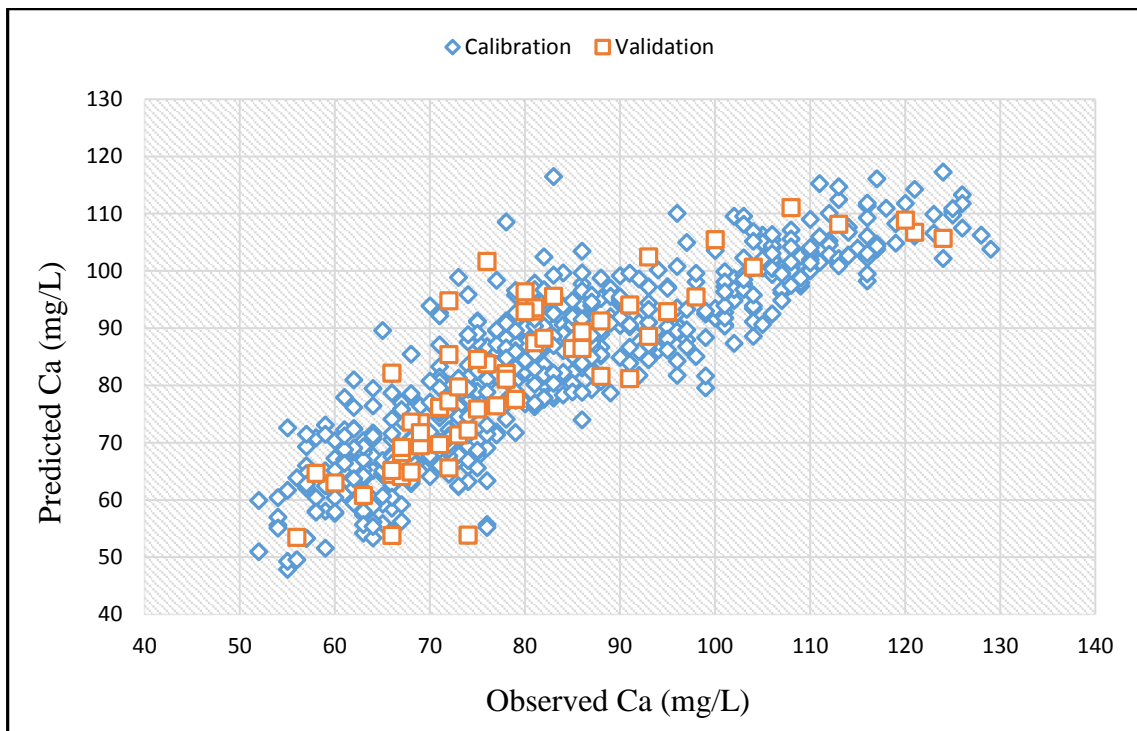


Figure 4. Ca-EC regression model.

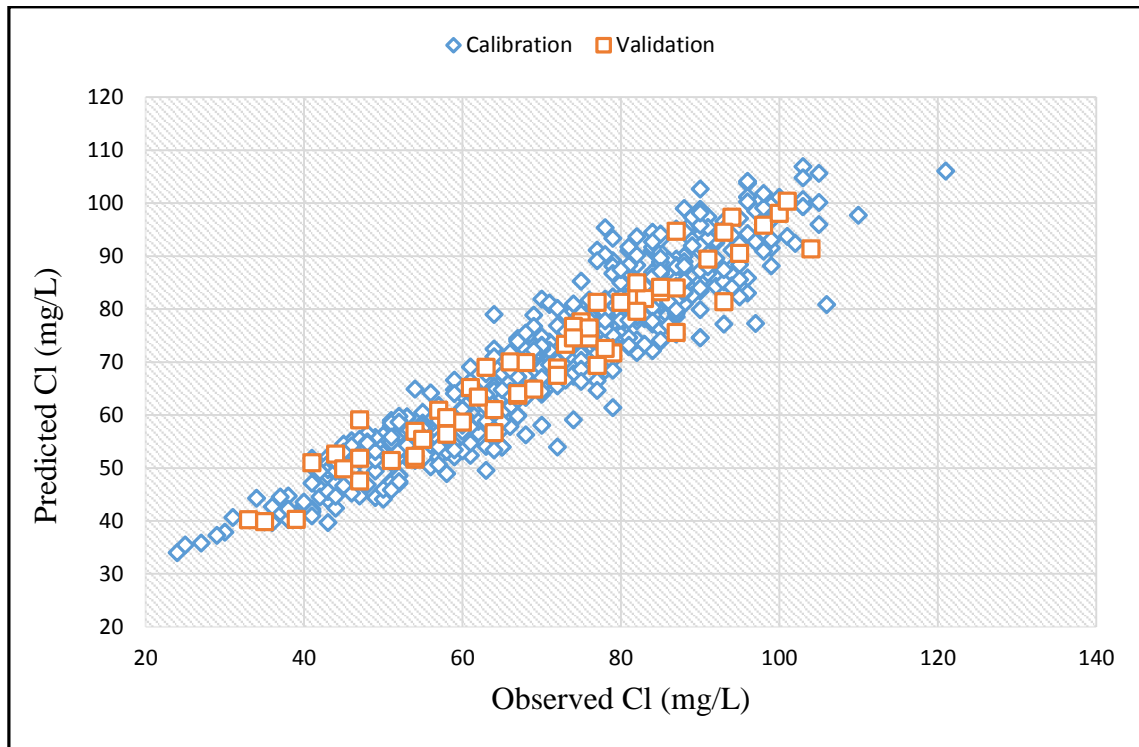


Figure 5. Cl-EC regression model.

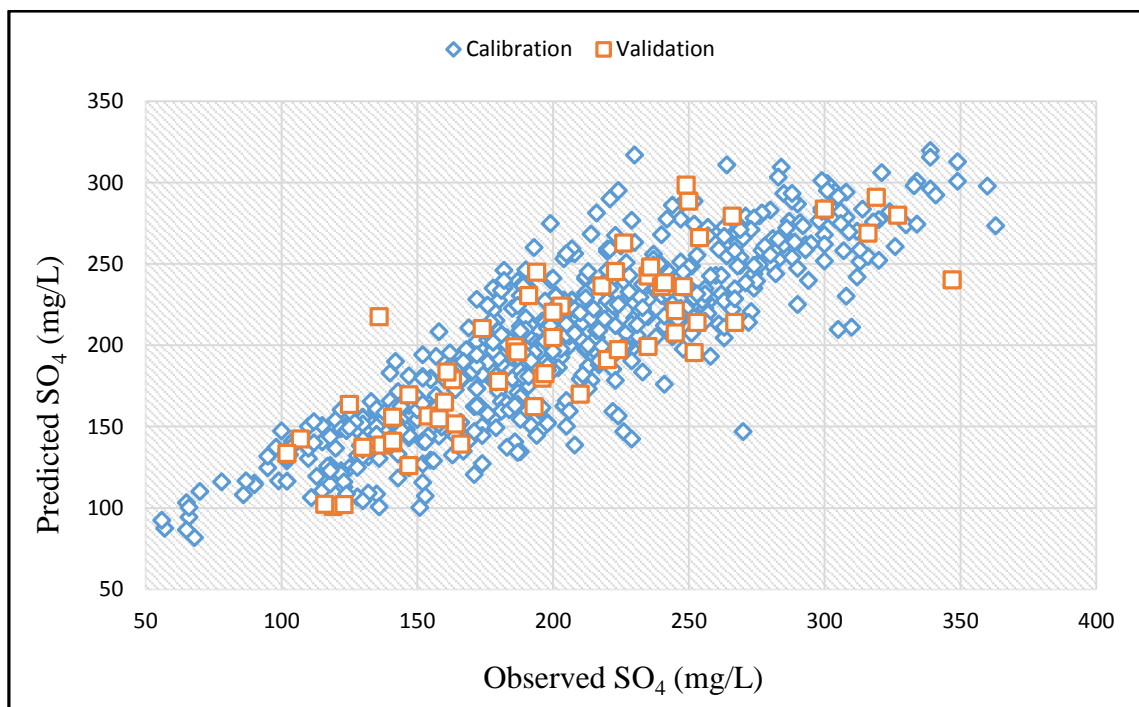


Figure 6. SO₄-EC regression model.



4. CONCLUSIONS

- It was concluded that the linear regression equations developed to predict the concentration of various parameters based on electrical conductivity can be used successfully within reasonable accuracy.
- EC can be used to predict various physicochemical parameters in the river's water.
- In the present study, five simple regression models were proposed showing a good/excellent prediction ability.
- The TDS-EC model is evident that the model is very accurate; while, the SO₄-EC model is less accurate.
- These five linear regression equations are beneficial as they represent a thumb rule for laboratory workers and water/ wastewater plant operators.
- A similar linear regression technique can be applied in any other locations within Rivers.

5. REFERENCES

- Al Jawad, M. S., and Tariq, B. Z., 2019. Estimation of Cutoff Values by Using Regression Lines Method in Mishrif Reservoir/ Missan oil Fields. *Journal of Engineering*, 25(2), pp.,82-95.
- APHA, 2017. *Standard methods for the examination of water and wastewater*. American Public Health Association.
- Berthouex, P. M., and Brown, L. C., 2002. *Statistics For Environmental Engineers*. CRC Press; 2nd Edition. p:512, ISBN-13 : 978-1566705929.
- Bhaduri, A., Bogardi, J., Siddiqi, A., Voigt, H., Vörösmarty, C., Pahl-Wostl, C., Bunn, S. E., Shrivastava, P., Lawford, R., and Foster, S., 2016. Achieving sustainable development goals from a water perspective. *Frontiers in Environmental Science*, Frontiers, 4, p. 64.
- Bhandari, N. S., and Nayal, K., 2008. Correlation study on physicochemical parameters and quality assessment of Kosi river water, Uttarakhand. *Journal of Chemistry*. Hindawi Publishing Corporation, 5(2), pp. 342–346.
- Joarder, M. A. M., Raihan, F., Alam, J. B., and Hasanuzzaman, S., 2008. Regression Analysis of Ground Water Quality Data of Sunamganj District, Bangladesh. *Int. J. Environ. Res.*, 2(3), pp., 291-296.
- Khudair, B. H., 2013. Assessment of Water Quality Index and Water Suitability of the Tigris River for drinking water within Baghdad City, Iraq. *Journal of Engineering*, 19(6), pp.,764-774.
- Krishnan, V., 2006. *Probability and Random Processes*. Wiley-Interscience; 1st Edition. p:723, ISBN-13 : 978-0471703549.
- Kumar, N., and Sinha, D. K., 2010. An approach to river water quality management through correlation study among various water quality parameters. *Int J Environ Sci*, 1(2), pp. 253–259.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W., 2005. *Applied linear statistical models*. McGraw-Hill Irwin New York.
- Peterson, J., McFarland, M., Dictson, N., Boellstorff, D., Berg, M., and Roberts, G., 2013. Texas Watershed Steward Handbook: A Water Resource Training Curriculum, Chapter 3: Overview of Watershed Impairments. The Texas Watershed Steward Program, *Texas AgriLife Extension Service*, Texas, United States, p:20, Publication: B-6203.



- Rusydi, A. F., 2018. Correlation between conductivity and total dissolved solid in various types of water: A review. in *IOP Conference Series: Earth and Environmental Science*. IOP Publishing, p. 12019.
- Snee, R. D., 1977. Methods and Examples Validation of Regression Models: Methods and Examples. *Technometrics*. Taylor & Francis Group, 19(4), pp. 415–428.
- WHO, World Health Organization, 2011. *Guidelines for Drinking Water Quality*. Fourth Edition. Vol.1, p:541, Recommendations, Geneva, Switzerland, ISBN: 978 92 4 154815 1.
- Zou, K. H., Tuncali, K., and Silverman, S. G., 2003. Correlation and simple linear Regression. *Radiology*. Radiological Society of North America, 227(3), pp. 617–628.