Statistical Remote Sensing for Prediction of Inland Water Quality Parameters for Shatt Al-Arab River in Iraq

Malik R. Abbas*1,2, Abd Wahid Bin Rasib3, Baharin Bin Ahmad1, Talib R. Abbas4

1Department of Remote sensing, Faculty of Built Environment and Surveying, university Teknologi, Malaysia, UTM, 81310 Johor Bahru, Johor, Malaysia
2Space & Communication Directorate, Ministry of Science and Technology, Baghdad-Iraq.
3Tropicalmap Map Research Group, Program geoinformation, faculty of Built Environment and Survey, university Teknologi malaysia, Skudai, 81310, Johor
4Environment and Water Directorate, Ministry of Science and Technology, Baghdad, Iraq

*Corresponding Author: maliksafaar@yahoo.com

Abstract. In this study, the empirical approach besides the methods multilinear regression was followed to generate the predictive mathematical models to estimate the water quality parameters values in the surface water of the Shatt Al-arab River which is located southern part of Iraq. And to show both benefits and the viability of using Landsat 8 optical spectral images to estimate some of the water quality parameters concentration. The daily water quality data archive for the of the total dissolved solids (TDS), conductivity (E.C), Nitrate (NO₃), and potential hydrogen ions (pH) of the water in four seasons (winter, spring, summer, and autumn) distributed along three years (2013, 2014, and 2015), was collected from four ground stations along of the Shatt Al-arab River. The objective of the study was to generate seasonal empirical mathematical models for the (open time) that can be used every year, without the need for calibration every time. Optical data were corrected to remove radiometric and atmospheric error sources effects prior to the developing the models. Multiple regression analysis between measured water quality parameters of the ground stations and the reflectance of the pixels corresponding to the sampling stations was used to generate these models. Determination coefficients (R²) of the proposed mathematical models were between 0.83-0.99. The percentage error between predicted and measured values for these models were between 0.03% -12%. The results of this work indicate the novelty of the approach used to generated these mathematical models for the open time for any year but in the each season. These models are reliable to estimate the spatial and temporal variation of TDS, E.C, NO₃, and pH. So models generated from Landsat 8 can be used as a tool to facilitate the environmental, economic, and social management of the surface waters bodies like a Shatt Al-arab River.

Keywords: Water Resources, Water Quality, Shatt Al-arab, Remote Sensing modeling.

1. Introduction
The topic of water quality is concerning with a major environmental issue as it impacts human life. (He et al., 2008 and Chen et al., 2009). One of the significant factors that have an influence on the...
physical dimensions (such as color) of the water is the characteristics (distribution and concentration) of fine suspended solids and dissolved matters (Clark et al., 1980). Many previous works studied the Parameters that impact water quality through the use of satellite data.

Using the methodology of remote sensing numerous works have explored the relationship between the parameters of water quality and radiance or reflection. Using the methodology of remote sensing, Multiple parameters, including physico-chemical parameters TDS, TSS, total alkalinity, temperature, turbidity, EC, salinity, as well as organic parameters (biochemical oxygen demand, DOC, BOD, TIC, TOC) and microbiological parameters (total coliform TC, Chlorophyll Chl-a) have been studied to be sensed remotely. Commonly, these water quality parameters can be correlated with the radiance or reflection from water surface. These correlations are normally vary seasonally. Therefore, any research dealing with the aforementioned relationship must be addressed to a specific season and conditions. (Abdelmalik, 2018).

Detailed description of remotely sensed data which can be implemented in water quality studies were presented by many authors as Kutser (2009), Matthews (2011) and Gholizadeh et al. (2016). Based on their spatial resolution these authors categorize this additional Data into three main classes. These classes are generally: first a high-resolution data such as Worldview Series data, Quick bird, and IKONOS. Second Moderate-resolution data: just like EO-1 Hyperion, ALI about 10–30 m, ASTER is ranged 15–90 m as well as Landsat series 15–120 m. The third class is Regional/Global resolution data: like terra MODIS 250-1000 m, NOAA-16 AVHRR 1000 m, SeaWiFS 1130 m, and ATSR-1, 2 (1000 m).

Compared to remote sensing approach the ground stations system or on-site surveys of detecting and investigating water quality variables of water are costly and need relatively much time, especially if they deal with vast areas. Remote sensing via satellites for the evaluation of water quality provides a relatively easy way to do the job without the need for costly on-site surveys (Hadjimitsis et al., 2010). Variables governing water quality making use of satellite techniques have been studied by many authors. The techniques of multiple-regression were applied to assess water quality parameters as Choubey and Subramanian (1992), Coskun et al. (2001), Hirthle and Renz (2003), Guan (2009), and Khattab and Merkel (2014).

The most commonly used models for surface water bodies are the empirical models. These models are usually developed by trying statistical correlations that relate the ground station data of water quality parameters and satellite data of spectral values. These models often present reliable site-specific relations. However, because of site-specific nature of the constituents of water bodies the accuracy of these empirical models decreases when applied to other water bodies (Chang et al., 2015). The goal of the present study was to find the best algorithms to assess pH, EC, TDS, Total alkalinity, Orthophosphorous for the Shatt Al-Arab River using the statistical methodology and Landsat 8 Data.

2. **Study area**

The river of Shatt Al-Arab is confluence result of combination of the Tigris and the Euphrates Rivers at Qurna town in the Basra Governorate of southern Iraq (Fig 1). The southern part of Shatt Al-Arab is formed the border between Iraq and Iran then it discharges into the Arabian Gulf with a total length of about 192 km (Abbas et al., 2015; Abbas et al., 2020). It varies in width from about 232 meters (761 ft) at Basra to 800 meters (2,600 ft) at its mouth. Water hydrology of many branches connected to this river are usually affected by tide and ebb cases due to wedge intrusion from the Arabian Gulf. The water of Shatt al-Arab carries so large amounts of silt that the river has to be frequently dredged to remain navigable. The two major cities of the river are Abadan in Iran and Basra in Iraq (Hamdan et al., 2018). Where the geographical location of Shatt al-Arab lies between latitudes 29.58 to 31.00 N and between longitudes 47.25 to 48.30 E.
3. Materials and methods

In this study, Landsat satellite data and water quality samples were used to derive general algorithms that are accurate for retrieving water quality parameters for the Shatt Al-Arab River using Landsat 8 reflectance data. The general flowchart describing the adopted methodology of this research. These include images pre-processing, water quality parameters estimations, as displayed in Figure 2.

3.1 Data collection and water quality parameters

Water quality data archive of daily values for the of total dissolved solids (TDS), electric conductivity (EC), Nitrate (NO₃), and potential hydrogen ions (pH) for four seasons (winter, spring, summer, and autumn) for three years 2013, 2014, and 2015 were collected from four ground stations along of Shatt Al-arab River. They start from the location of Qurna town north of Basra City, it is the confluence of Tigris and Euphrates River. They are SH1 SH2 SH3 and SH4. (Fig.3). These stations were selected to carry out the present study along the 170 km stretch of Shatt Al-arab River situated along the path from north to the south of Basra Governorate. The data used in this paper were provided by the Iraqi Ministry of Environment, as well as were analyzed in the laboratory using the standard procedures (Federation, 2005).

3.2 Landsat 8 image processing

To develop better and more accurate statistical relationship between satellite image data and water quality parameters sampling is better to be at the same days of satellite images date (Bonansea et al. 2015; Brezonik et al. 2005; Kloiber et al. 2002). Total number of images of Landsat-8 used in this paper was 20 images type GeoTIFF level 1 (path 166; row 39) distributed among four seasons along four years. The images were downloaded from the US Geological Survey (USGS) database using the Global Visualization Viewer (http://glovis.usgs.gov/).

There are two main sensors borne on satellite Landsat 8. The first sensor is the multispectral sensor Operational Land Imager (OLI) and the other sensor is the Thermal Infrared Sensor (TIRS). (Roy et al. 2014). These sensors are sensing eleven spectral bands in which are B1-B7 with 30 m resolution (coastal/aerosol, blue, green, red, near infrared SWIR1, SWIR2, respectively) in addition to B9 (cirrus), B8 (panchromatic) with a resolution of 15 m, and B10 and B11 (TIR-1 and TIR-2) with a resolution of 100 m. To remove atmospheric effects and to transform the relative values of pixels data
into absolute measurements. Radiometric and atmospheric corrections were applied to the images. The images were processed in GIS (version 10.5).

Figure 2: Flowchart of research methodology.
3.3 Statistical relations to estimate water quality parameters
To obtain statistical relationships between ground stations measurements and Landsat 8 images data different mathematical combinations of reflectance bands (individual, square roots, reciprocals, square, and band ratios, etc) of the processed Landsat images were considered as independent variables while the ground stations measurements data were used as dependent variables with the aid of Matlab version 2015. The statistical relations were developed using the successive steps multiple regression technique. The accuracy of the developed statistical relations was evaluated in terms of root means square error (RMSE) and the coefficient of determination ($R^2$).

3.3.1 Correlation Coefficient and Multi Linear Regression. Water quality is normally characterized by many physico-chemical parameters. Due to many influencing factors these parameters change widely. These factors may include source of water, type of pollutants, seasonal fluctuations, etc. Statistical analysis such as correlation and regression analysis of the physico-chemical properties of a surface water give an accepted amount of information like their average values and possibly prediction of one variable (usually the one which is difficult to evaluate). In general correlation coefficient represents the associated way of variation between two variables under consideration. It provides a measure where these two variables increase and decrease simultaneously in harmonic manner or whether one variable decreases as the second increases in harmonic manner. When their patterns of variation are totally unrelated that means no correlation is exists between these two variables. So, the correlation measures no more than the degree of co-variation between the two variable. Correlation coefficient between two variable does not provide evidence for causal relationship between them. However, one may cause the other, as NO$_3$ high concentration causes high algae concentration, or both share the same cause, such as two heavy metals concentration measured at a variety of times at a variety of locations. The understanding of the process involved is the only way to get evidence for causation. Correlation coefficient is dimensionless and scaled to lie in the range $-1 < r < 1$. When its value is zero, $r = 0$, that is mean there is no correlation between the two variables. When one variable increases as the second increases, ‘r’ is positive. When they vary in opposite directions, ‘r’ is negative.
Correlation becomes a test for temporal trend when one variable is a measure of time while it becomes test for spatial trend when one variable is location. Variables may be correlated in either a linear or nonlinear manner. When one variable generally increases or decreases as the other variable increases or decreases, the two variables are said to possess a monotonic correlation. This correlation may be nonlinear, with exponential patterns, piecewise linear patterns, or patterns similar to power functions when both variables are non-negative. This non-linearity indicates that a measure of linear correlation would be inappropriate. (Berthouex and Brown 2002).

4. Results and discussion
Using linear regression technique of successive steps, satisfactory statistical models were developed to estimate of total dissolved solids (TDS), electric conductivity (EC), Nitrate (NO$_3$), and potential of hydrogen ions (pH) in Shatt Al-arab River water for the four season (winter, spring, summer and autumn). The generated models at (P < 0.05) are presented in Table 1.

| Parameters | Season | Mathematical formula | $R^2$ Model | $R^2$ Validation |
|------------|--------|----------------------|-------------|-----------------|
| **TDS**    | Winter | TDS = -23655138.1448246* (B5*B4) + (634372486.44534 * (B5*B4) + (768140984302.496 * (B5*B4) ^3) + (42897749915570.4 * (B5*B4) ^4) - (897061776138517 * (B5*B4) ^5) + (389887.46936319 * (B4+B5) ^2) + (-1351785.90427472 * (B4+B5) ^3) + 3076.48756132925 | 0.95 | 0.86 |
|           | Spring | TDS = (531611.3899 * (B5*B4)) + (-1794142245 * (B5*B4) ^3) + (66895023675 * (B5*B4) ^4) + (-6.91226*11 * (B5*B4) ^5) + (-297227.0033 * (B4+B5)) + (1091942.689 * (B4+B5) ^2) + (-1323059.993 * (B4+B5) ^3) + 23668.45291 | 0.95 | 0.96 |
|           | Summer | TDS = (-965117.89823314 * (B5*B3)) + (61836952.9125505 * (B5*B3) ^2) + (-206344899.543716 * (B5*B3) ^3) + (37856129.110334 * (B5*B3) ^4) + (-362835495.030328 * (B5*B3) ^5) + (142164628.653782 * (B5*B3) ^6) + 612187.76672194 | 0.93 | 0.94 |
|           | Autumn | TDS = (-264305.6956 + (-1483666472 * (B5*B4) ^2) + (-35036636255 * (B5*B4) ^3) + (-443382887839.433 * (B5-B4) ^4) + (-2867511811.14294 * (B5-B4) ^5) + (-744213901640.21 * (B5-B4) ^6) + (3169968.92 * (B4-B5)) | 0.91 | 0.95 |
| **EC**    | Winter | E.C = (-17210.2727 + (120686101.5 * (B5) ^2) + (-5584779398.57767 * (B5) ^3) + (10181027295.578 * (B5) ^4) + (829710187463.317 * (B5) ^5) + (25128329986568.82 * (B5) ^6)) | 0.99 | 0.89 |
|           | Spring | E.C = (-296480.8198 + (19522582.17 * (B4+B2)) + (-166070597.2 * (B4+B2) ^2) + (734010303.9 * (B4+B2) ^3) + (-1782125792.217 * (B4+B2) ^4) + (2258340472.456 * (B4+B2) ^5) + (-1169022917.531 * (B4+B2) ^6)) | 0.93 | 0.94 |
|           | Summer | E.C = (298568914.2 + (-16332721873.3022 * (B4)) + (370225581112.511 * (B4) ^2) + (-4451453804040.34 * (B4) ^3) + (29944801820266.6 * (B4) ^4) + (-106865767185936 * (B4) ^5) + (158083019911297 * (B4) ^6)) | 0.97 | 0.96 |
The water quality models developed in this study show the viability of Landsat 8 images application in the characterization of total dissolved solids (TDS), electric conductivity (EC), Nitrate (NO$_3$), and potential of hydrogen ions (pH). These models provide the ability to estimate the water quality parameters distribution along the Shatt Al-Arab River in the south of the Iraq area instead of only at the sampling stations. That is the first benefit of these mathematical models. Another benefit, which is more important, they can be applied for the same season at different years. The results based on R$^2$, RMSE, standard error (SE), and (P-value) values (Table 2) show the viability of these types of models for application in the area of south Iraq regarding the Shatt Al-Arab River only.

The four bands (B2, B3, B4, and B5) of Landsat 8 contributed to the generation of water quality models. Starting from the simplest models. EC models generated for winter, summer, and autumn season which was generated with a single band (B5) for winter and (B4) for the summer and autumn season. Also for NO$_3$ models with a single band, B4, B3 and B2 for winter, spring, and summer respectively. For pH models were a single band for all seasons B4, B5, B4, and B5 respectively for winter, spring, summer, and autumn. And for the other parameter like TDS was too more complex models in this study. It was generated by combined of bands B3, B4, and B5 as well as it's the bands' ratios. However, the models generated in the present study differ from reported models in the scientific literature.

Table 2: Regression accuracy information of proposed models of TDS, EC, NO$_3$, and pH generated from Landsat 8 images for the Shatt Al-Arab River water during four seasons.
Usually, one of the main drawbacks of the published empirical models is that they need a recurring parameterization and calibration due to the changing nature of the constituents of surface water bodies (Yacobi et al., 2011; Chang et al., 2014). However, all models in the literature were created as simple models, used a simple correlation for data based on one small area as well as for a special month of the year. In the current study by following different approaches in collecting data and special statistical processing methods, including data collecting in the months but for different years, the seasonal generated mathematical models can be used for the open time for any year for the specific season. It was a novelty approach. The implementation of a hybrid system composed of traditional monitoring system and the application of satellite remote sensing for the estimation of parameters of water quality, could enhance the development of the environmental, economic, and social management of the river.

### 4.2 Proposed models Verification

For the purpose of validation, graphically the relationship between values water quality parameters estimated through the models was compared with values water quality parameters on-site by the ground stations. Figures 4 to 19 shows the validations of models results for each water quality parameters used in this current study TDS, EC, NO₃ and pH respectively. Generally, the results displayed a significant correlation between the measured values and the predicted values measured values. Each figure have shown a three parts of information. The first part is for the model results and the second part is for the validation of the model's results, in addition the third part is for shown the percentage error of the proposed model and its percentage error of the validation result.
Figure 4: Comparison of the values of the TDS parameter of the model, model's validation and percentage error of the model for the winter season of Shatt Al-arab River.

Figure 5: Comparison of the values of the TDS parameter of the model, model's validation and percentage error of the model for the spring season of Shatt Al-arab River.
Figures 4 to 7 show the predictive values of the TDS parameter that are got by the proposed mathematical model for seasons winter, spring, summer, and autumn respectively, compared with the actual values of TDS by the ground stations that are related to the Shatt Al-arab River. These figures show the average of a percentage error for the proposed models of the TDS parameter which has been ranged from 6.2, 2, 3, and 5.3 for four seasons winter, spring, summer, and autumn respectively. As well as percentage error has been ranged 8, 6, 2, and 3 for four seasons winter, spring, summer, and autumn respectively.
autumn respectively which regard the validation results. The presented results showed that the models have high accuracy. With the same analysis for the other Figures 8-19, of the water quality parameters E.C, NO₃, and the pH respectively of the water of Shatt Al-arab River.

Figure 8: Comparison of the values of the E.C parameter of the model, model's validation and percentage error of the model for the winter season of Shatt Al-arab River.

Figure 9: Comparison of the values of the E.C parameter of the model, model's validation and percentage error of the model for the spring season of Shatt Al-arab River.
Figure 10: Comparison of the values of the E.C parameter of the model, model's validation and percentage error of the model for the summer season of Shatt Al-arab River.

Figure 11: Comparison of the values of the E.C parameter of the model, model's validation and percentage error of the model for the autumn season of Shatt Al-arab River.
Figure 12: Comparison of the values of the NO$_3$ parameter of the model, model's validation and percentage error of the model for the winter season of Shatt Al-arab River.

Figure 13: Comparison of the values of the NO$_3$ parameter of the model, model's validation and percentage error of the model for the spring season of Shatt Al-arab River.
Figure 14: Comparison of the values of the NO$_3$ parameter of the model, model's validation and percentage error of the model for the summer season of Shatt Al-arab River.

Figure 15: Comparison of the values of the NO$_3$ parameter of the model, model's validation and percentage error of the model for the autumn season of Shatt Al-arab River.
Figure 16: Comparison of the values of the pH parameter of the model, model's validation and percentage error of the model for the winter season of Shatt Al-arab River.

Figure 17: Comparison of the values of the pH parameter of the model, model's validation and percentage error of the model for the spring season of Shatt Al-arab River.
Figure 18: Comparison of the values of the pH parameter of the model, model's validation and percentage error of the model for the summer season of Shatt Al-arab River.

Figure 19: Comparison of the values of the pH parameter of the model, model's validation and percentage error of the model for the autumn season of Shatt Al-arab River.
5. Conclusions
The results of the current study indicates that there exists an empirical relationship between the in situ water quality values of the Shatt Al-arab River and the optical remote sensing data Landsat 8. A multilinear regression approach is used for establishing the relationship between water surface reflectance and water quality parameters concentration that are collected in the temporary observatories which along the river sites. Final output results of the proposed mathematical models allow the interpretation of spatial patterns of TDS, E.C, NO$_3$, and pH in the study area entirely, and the conditions of the water quality conditions appear to agree with the expected pollution sources in the river sites. The estimated results obtained from Landsat 8 have a close relationship with ground measured data. The proposed mathematical models are used to predictive the distributions of concentrations of water quality parameters in the surface water bodies. It's helpful in making decisions regarding the improvement of urban water management as well as the identification of potential problems in water ecosystems and human health. Therefore, the results of this study show the feasibility of using Landsat data for monitoring water quality parameters concentrations of small-sized or narrow-width urban water bodies. The analysis results indicate the potentials of remote sensing in initiating a cheap and effective method for monitoring polluted rivers in Iraq on a routine basis. Additionally, this study presents the technical potentials of Landsat data in estimating the concentration of water quality parameters in urban water bodies. As well as provide a scientific basis for inquiring the means for the extension of a limited set of field observations to times or areas where field data are unavailable. There is a need for further studies for the examination of pollutant determination from Landsat data over the study area and other urban water bodies.

Acknowledgment
The authors would like to express their profound gratitude to the Faculty of Built Environment and Surveying (FABU), Universiti Teknologi Malaysia (UTM) for all the support that been provided. The authors also would like to thank UTM for providing financial support through UTM CR DTD VOT 4C255 and UTM IIIG VOT 01M78.

References
Abdelmalik, K. W. (2018). Role of statistical remote sensing for Inland water quality parameters prediction. *The Egyptian Journal of Remote Sensing and Space Science, 21*(2), 193-200.

Hadjimitsis, D. G., Hadjimitsis, M. G., Toulios, L., & Clayton, C. (2010). Use of space technology for assisting water quality assessment and monitoring of inland water bodies. *Physics and Chemistry of the Earth, Parts A/B/C, 35*(1-2), 115-120.

Chang, N. B., Imen, S., & Vannah, B. (2015). Remote sensing for monitoring surface water quality status and ecosystem state in relation to the nutrient cycle: a 40-year perspective. *Critical Reviews in Environmental Science and Technology, 45*(2), 101-166.

Hamdan, A., Dawood, A., & Naeem, D. (2018). Assessment study of water quality index (WQI) for Shatt Al-arab River and its branches, Iraq. In *MATEC Web of Conferences* (Vol. 162, p. 05005). EDP Sciences.

He, W., Chen, S., Liu, X., & Chen, J. (2008). Water quality monitoring in a slightly-polluted inland water body through remote sensing—Case study of the Guanting Reservoir in Beijing, China. *Frontiers of Environmental Science & Engineering in China, 2*(2), 163-171.

Chen, S., Fang, L., Zhang, L., & Huang, W. (2009). Remote sensing of turbidity in seawater intrusion reaches of Pearl River Estuary—A case study in Modaomen water way, China. *Estuarine, Coastal and Shelf Science, 82*(1), 119-127.

Clark, D. K., Baker, E. T., & Strong, A. E. (1980). Upwelled spectral radiance distribution in relation to particulate matter in sea water. *Boundary-Layer Meteorology, 18*(3), 287-298.
Matthews, M. W. (2011). A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. *International Journal of Remote Sensing*, 32(21), 6855-6899.

Kutser, T. (2009). Passive optical remote sensing of cyanobacteria and other intense phytoplankton blooms in coastal and inland waters. *International Journal of Remote Sensing*, 30(17), 4401-4425.

Khattab, M. F., & Merkel, B. J. (2014). Application of Landsat 5 and Landsat 7 images data for water quality mapping in Mosul Dam Lake, Northern Iraq. *Arabian Journal of Geosciences*, 7(9), 3557-3573.

Guan, X. (2009). *Monitoring Lake Simcoe water quality using Landsat TM images* (Master's thesis, University of Waterloo).

Hirtle, H., & Rencz, A. (2003). The relation between spectral reflectance and dissolved organic carbon in lake water: Kejimkujik National Park, Nova Scotia, Canada. *International Journal of Remote Sensing*, 24(5), 953-967.

Coskun, G., Ekercin, S. O., & Ztopal, A. (2001, May). Monitoring and water pollution modelling of the Bosphorus by regression analysis using Landsat-TM data. In *EARSeL 21th Symposium Observing our Environment from Space/New Solutions for a New Millennium, ENSG* (pp. 14-16).

Choubey, V. K., & Subramanian, V. (1992). Estimation of suspended solids using Indian Remote Sensing Satellite-IA data: a case study from Central India. *International Journal of Remote Sensing*, 13(8), 1473-1486.

Gholizadeh, M. H., Melesse, A. M., & Reddi, L. (2016). A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors*, 16(8), 1298.

Federation, W. E., & American Public Health Association. (2005). Standard methods for the examination of water and wastewater. *American Public Health Association (APHA): Washington, DC, USA*.

Abbas, M. R., Ahmad, B. B., & Abbas, T. R. (2015). Use MODIS satellite data to study new phenomena of underground fire in the Al Ruhban oasis in Al Najaf city, Iraq. *Environmental Earth Sciences*, 73(7), 3475-3485.

Abbas, M. R., Hason, M. M., Ahmad, B. B., & Abbas, T. R. (2020). Surface roughness distribution map for Iraq using satellite data and GIS techniques. *Arabian Journal of Geosciences*, 13(17), 1-13.