Journal homepage www.jzs.univsul.edu.iq
Journal of Zankoy Sulaimani
Part-A- (Pure and Applied Sciences)

Water Quality Assessment Models for Dokan Lake Using Landsat 8 OLI Satellite Images

Hasti Shwan Abdullah[1], Mahmoud S. Mahdi[2] & Hekmat M. Ibrahim[1]

1Faculty of Engineering-Sulaimani University, Bakrajo Street, Sulaimaniyah-Iraq
2Building & Construction Engineering Department-University of Technology, 52 Street, Baghdad- Iraq.
Email: hekmat.ibrahim@univsul.edu.iq

Article info

Original: 15 November 2016
Revised: 15 April 2017
Accepted: 16 May 2016
Published online: 20 September 2017

Key Words:
Dokan Lake
WQPs
Image Processing
Landsat 8
GIS

Abstract

It is impractical to monitor water quality more than a small fraction of lakes by conventional field methods because of expense and time requirements. Satellite image is more convenient to be applied to collect the required data for monitoring and assessing water quality in the lakes. Therefore, this study aims to estimate the concentration of some water quality parameters (Temperature, DO, BOD, pH, Turbidity, TSS, TDS, EC, NO_3, PO_4 and E. coli) by applying developed models based on the remote sensing and GIS techniques on the Landsat 8 OLI satellite image using twenty points in Dokan lake, Kurdistan Region, Iraq at two different seasons.

Multiple linear regression is used to obtain mathematical models for estimating the concentration of some water quality parameters depending on spectral reflectance of Landsat 8 OLI. In this study, new band (coastal blue) of Landsat 8 OLI has been undertaken in developing of models. Moreover, new Independent Component Analysis (ICA) and new 7 band ratios with 16 band combinations have been used. The best models are obtained for TSS, Turbidity and DO with coefficient of determination (R^2) of 0.98, 0.98, and 0.83 respectively. Generally, for spring season, the performance of all models is reduced due to seasonal change, variance of parameters and other factors. However, high R^2 of 0.86 has been shown for Temperature.

The results of the developed WQPs models have been mapped to show the water quality parameters concentration distribution within Dokan lake. The conclusions present that correlation of all bands of Landsat 8 OLI is appropriate to water quality parameters.

Introduction

Water quality monitoring is the systematic collection and evaluation of data about the chemical, physical, and biological quality of the water bodies, and assesses how external changes, both natural and anthropogenic, affect that quality.

To get a true picture about the nature of the river and lake water, it is required to measure the quantity and quality of water through water quality monitoring, which implemented in many methods and techniques. The traditional method of water quality monitoring consists of collecting and analysing water samples after testing them in laboratory which requires more times and costs. Recently, with advance and increasing role of technology, new techniques and methods are developed for assessing water quality such as remote sensing (RS) and geographical information system (GIS) that achieved through using satellite data to monitor water quality to reduce time and cost for the process, and to increase accuracy of results. Ming, et al. [1] established an integrated water quality monitoring system data which obtained from SPOT data with GIS techniques in central Taiwan. Also, Hedger, et al. [2] highlighted the role that remote sensing may play in helping to optimize sampling that are more commonly implemented.

Remote sensing and GIS have a potential to monitor spatial variation in water quality over large areas. Doxaran, et al. [3] obtained linear relationships with high correlation between the concentrations of coloured dissolved organic and suspended (total, organic and inorganic) matter and remote-sensing Reflectance (Rrs) measurements in the Tamar estuary (south-west UK). Remote sensing of lake water is often limited to high spatial resolution satellites such as Landsat, which have limited spectral resolution. The first four bands of Landsat 7-ETM satellite data were correlated by Alparslan, et al. [4] with chlorophyll-a, suspended solid matter, Secchi disk transparency and total phosphate to assess water quality at Omerli dam, Istanbul city, Turkey.

Many studies have used the regression technique to correlate the water quality parameters with satellite image data. By multiple regression, ordinary kriging interpolation technique and aid of software packages, Nas, et al. [5] used the Terra ASTER satellite image as remote sensing data source for water quality parameter mapping in Beysehir lake, Central Anatolia, Turkey.

By increasing the number of satellites and developing their abilities in remote sensing, many researchers have investigated and compared water quality monitoring using different satellite images. Li, et al. [6] focused on the substantial differences between instruments on-board of Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and Landsat-8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) which are currently operating for routine earth observation.

Knowledge of the spatial distribution of different biological, chemical and physical variables is essential in environmental water studies as well as for resource management [7]. Therefore, this study presents the results of an investigation into satellite monitoring of lake water quality to ascertain the feasibility of estimating water quality parameters and its spatial distribution using Landsat 8 OLI imagery combined with in situ data from Dokan lakes which is one of the main sources of water for drinking, irrigation and power generation in Kurdistan Region and especially to the governorate of Sulaimani. The present study focuses on the modelling of more important water quality parameters using remote sensing and GIS. The study uses Landsat 8 OLI images to model some water quality parameters which includes pH, Total Dissolved Solids (TDS), Electrical Conductivity (EC), Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Nitrate (NO_3), Nitrate Nitrogen (NO_3-N), Phosphate (PO_4), Total Phosphorous (TP), Temperature, Total Suspended Solids (TSS) and Turbidity. A new band such as Coastal Blue of Landsat 8, which is not used before, and new ratios and combinations of spectral bands and their transforms such as Independent Component Analysis (ICA), Minimum Noise Fraction (MNF), are used in this study.

Methodology

Generally, the methodology of carrying out this study can be divided into two parts: theoretical and experimental. The theatrical part concerns the national and international historical background researches that used different techniques to find and evaluate the water quality parameters (WQPs) for local and international rivers and lakes and that related to the use of remote sensing and GIS techniques in this issue. The experimental part, on the other hand, concerns the field work, laboratory test and data analysis. Figure-1 shows the flow chart of carrying out the present study.

Materials and Methods

A. Study Area

Dokan lake is located in Iraq on the Lower Zab River approximately 295 km north of Baghdad and 65 km northwest of Sulaimani city as shown in Figure-2. It is surrounded by mountains of Sara and Quasar to the southeast, Assos to the northeast, Kosrat to the southwest and Barda Rash to the northwest [8].

The reservoir impounded by Dokan Dam which has a total design capacity of 6.870 billion cubic meters (6.140 billion cubic meters is live storage and 0.730 billion cubic meters being dead storage) at normal operating level of 511 m.a.s.l. The current storages are less than this due to over 50 years of sedimentation [9] as cited in [10].

B. Remote Sensing

The detection of electromagnetic energy can be performed either photographically or electronically. Imagery can be expressed in four dimensions, spectral, temporal, radiometric and spatial resolutions [11]. The general processes and elements involved in remote sensing of earth resources are grouped into two basic processes, data acquisition and data analysis processes [12].

Since 1972, Landsat satellites have continuously acquired space-based images of the Earth's land surface, coastal shallows, and coral reefs. Landsat 8 data have been used by government, commercial, industrial, civilian, military, and educational communities throughout the United States and worldwide [13]. Therefore, the Landsat 8 data are used in the present study to determine some water quality parameters concentrations of Dokan lake water.

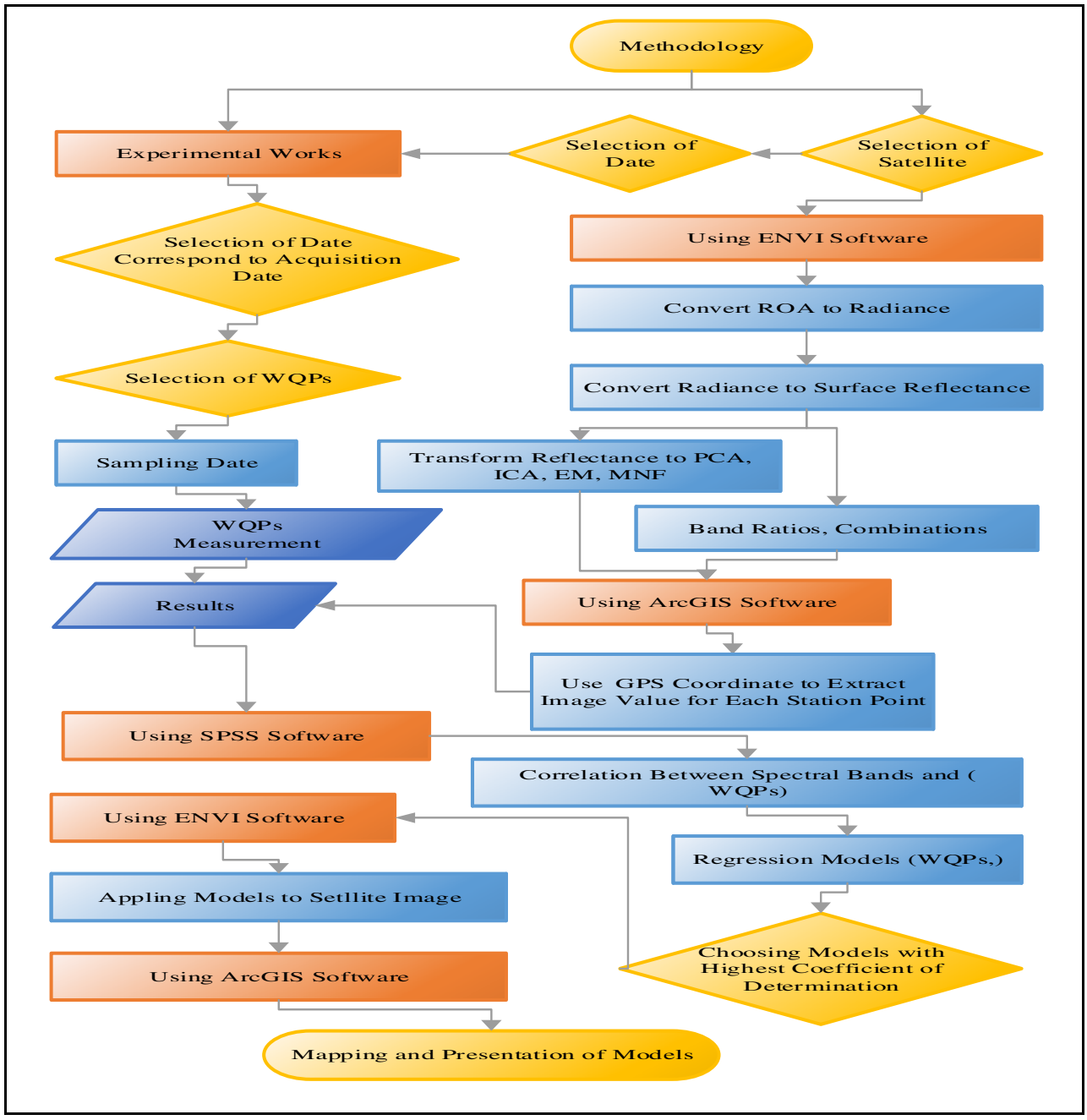

Figure-1: Flowchart of carrying out the methodology of the present study.

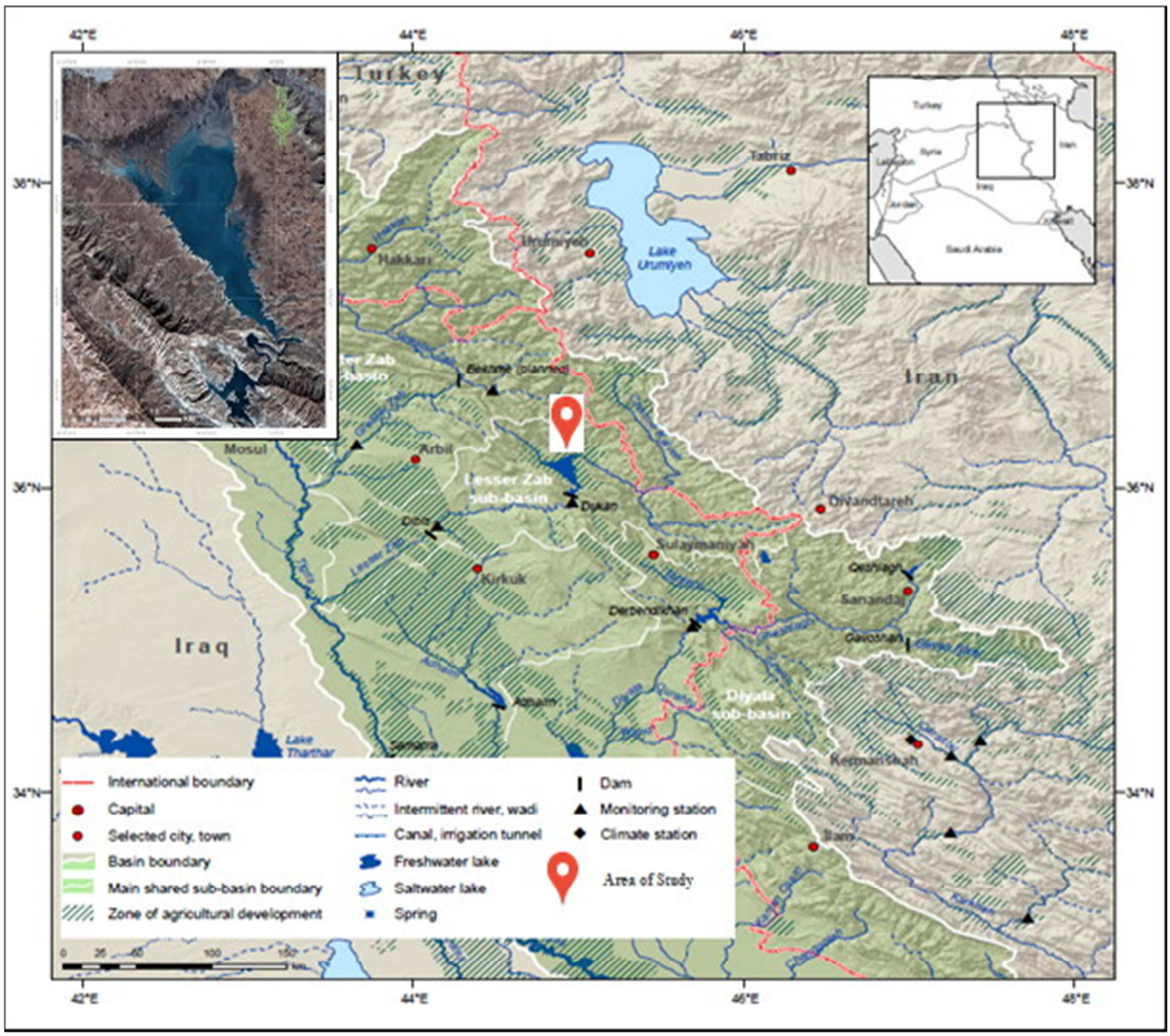

Figure-2: Map of Northern Iraq showing the study area as cited in [14].

C. Image Pre-processing

The satellite data image processes include two groups, the first is pre-processing and the second is image enhancement and transformations. The pre-processing commonly comprises a series of sequential operations, including geometric correction, atmospheric correction or image registration, normalization, masking (e.g., for clouds, water, irrelevant features), image rectification, and image re-sampling [15]. There are two formulas that can be used to convert Digital Numbers (DNs) to radiance [16]. One method uses the Gain and Bias values from the header file. The method converts DNs to at-sensor spectral radiance L, also called top-of-atmosphere radiance:

$$L_\lambda = Gain * QCal + bias \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

Where:

L_λ= Spectral Radiance at the sensor's aperture in watts/ (m^2 *ster*µm).

$Gain$ = Rescaled gain (the data product "gain" contained in the Level 1 product header or ancillary data record) in watts / (m^2*ster*µm).

$QCal$ = the quantized calibrated pixel value in DN.

$bias$ = Rescaled bias (the data product "offset" contained in the Level 1 product header or ancillary data record) in watts/(m^2*ster*µm).

The longer method uses the LMin and LMax spectral radiance scaling factors. Coefficients are provided in one of three band-specific formats: gain and offset; $Grescale$ (also called gain) and $Brescale$ (bias); or radiances associated with minimum and maximum DN values ($Lmax$ and $Lmin$). Any of the three can be used to convert from DN to at-sensor radiance:

$$L_\lambda = \frac{LMAX_\lambda - LMIN_\lambda}{QCALMAX - QCALMIN}(QCAL - QCALMIN) + LMIN \quad \ldots\ldots\ldots\ldots \quad (2)$$

Where:

L_λ= the cell value as radiance.

$QCAL$= digital number.

$LMIN_\lambda$= spectral radiance scales to QCALMIN.

$LMAX_\lambda$ = spectral radiance scales to QCALMAX.

$QCALMIN$ = the minimum quantized calibrated pixel value (typically = 1).

$QCALMAX$ = the maximum quantized calibrated pixel value (typically = 255).

New characteristics are added to enhance and automatize ground reflectance retrieval by conversion to top of atmosphere reflectance units [16] as follow:

$$\rho_\lambda = \frac{\pi L_\lambda d^2}{ESUN_\lambda sin\theta} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

Where:

L_λ= Radiance in units of W/ (m^2 * sr * µm).

d= Earth-sun distance, in astronomical units.

$ESUN_\lambda$= Solar irradiance in units of W/ (m^2 * µm).

θ= Sun elevation in degrees.

D. Image Processing

The objectives of the second group of image processing functions are grouped under the term of image enhancement and image transformations [17]. Some of image transformations that can be made by the ENVI software and used in this study are: Minimum Noise Fraction (MNF), Principal Component Analysis (PCA), Independent Component Analysis (ICA), band ratios and indices [18]. Table-1 shows some spectral indices that used in the present study.

Another process includes the image classification and analysis which are used to digitally identify and classify pixels in the data. Classification is usually performed on multi-channel data sets (A) and this process assigns each pixel in an image to a particular class or theme (B) based on statistical characteristics of the pixel brightness values [17].

Table-1: Some spectral indices that used in the present study.

No.	Spectral Index	Equation
1	Normalized Difference Vegetation Index (NDVI)	$\mathrm{NDVI} = (\mathrm{NIR} - \mathrm{R})/(\mathrm{NIR} + \mathrm{R})$
2	Modification of Normalised Difference Water Index (MNDWI)	$\mathrm{MNDWI} = (\mathrm{G} - \mathrm{SWIR1})/(\mathrm{G} - \mathrm{SWIR2})$
3	Land Surface Water Index (LSWI)	$\mathrm{LSWI} = (\mathrm{NIR} - \mathrm{SWIR1})/(\mathrm{NIR} - \mathrm{SWIR1})$
4	Normalized Burn Ratio (NBR)	$\mathrm{NBR} = (\mathrm{NIR} - \mathrm{SWIR2})/(\mathrm{NIR} + \mathrm{SWIR2})$
5	Moisture Stress Index (MSI)	$\mathrm{MSI} = (\mathrm{G} - \mathrm{NIR})/(\mathrm{G} + \mathrm{NIR})$
6	Normalized Difference Moisture Index (NDMI)	$\mathrm{NDMI} = (\mathrm{R} - \mathrm{NIR})/(\mathrm{R} + \mathrm{NIR})$
7	Ratio-Vegetation-Index (RVI)	$\mathrm{RVI} = (\mathrm{NIR})\ /(\mathrm{R})$
8	Infrared Percentage Vegetation Index (IPVI)	$\mathrm{IPVI} = (\mathrm{NIR})/(\mathrm{R} + \mathrm{NIR})$

The remote sensing data image which used in this study include two Landsat-8 OLI images (path: 169 and row: 35) acquired on 24 October 2014 as an (ID: LC81690352014297LGN00) and 02 April 2015 as an (ID: LC81690352015092LGN00). The images were downloaded from the USGS, Earth Explorer (EE)

website. Both images contained small pockets of clouds, which covered about 0.23% and 7.29% of the autumn and spring season images respectively. These values are less than the minimum removing percentage (9%) of full scene of satellite image USGS, 2015, so there is no need of haze of cloud removal. ENVI software has been used for geospatial analysis and spectral image processing of Landsat 8 OLI data imagery.

The images have been geometrically corrected and prepared for atmospheric correction which is carried out to minimize the atmospheric effects through converting DN values to radiance. To derive a radiance image from uncalibrated image, a gain and offset applied to the pixel values. These gain and offset values are typically retrieved from the image's metadata or received from the data provider. ENVI software provides a tool called radiometric calibration undertakes this process for many data products that are distributed with calibration gain and offset values in the metadata. Finally, radiance is converted to Surface Reflectance (SR) which represents the reflectance of the surface of the Earth as shown in Figure-3.

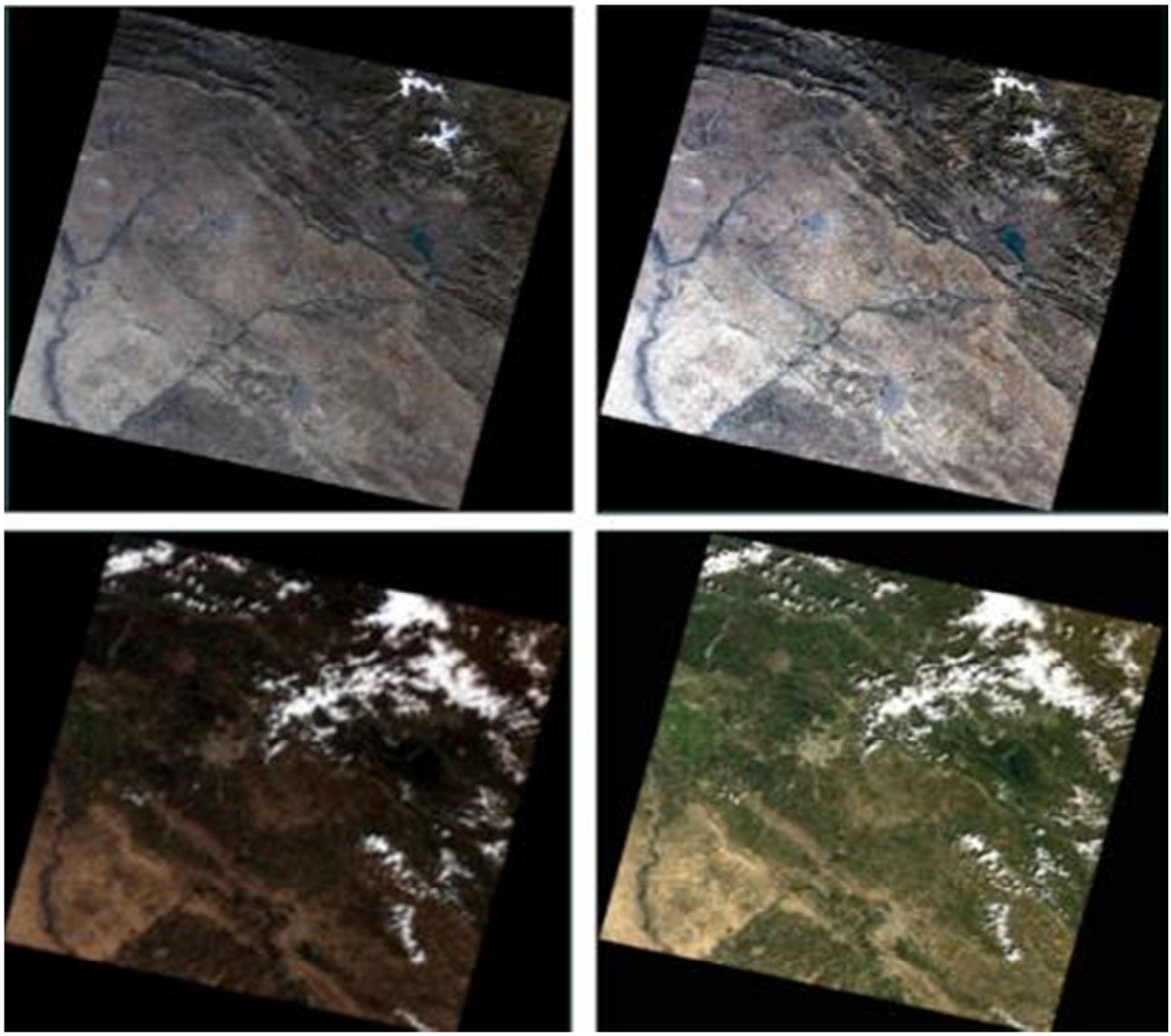

Figure-3: Landsat 8 OLI images of October 24th, 2014 (upper) and April 2nd, 2015 (lower), atmospheric and radiometric correction, (left) before, (right) after pre-processing.

Accordingly, the effects of clouds and other atmospheric components have been minimized. The SR reflectance is unlike the Top-of-Atmosphere (TOA) reflectance which is the reflectance measured by a space-based sensor flying higher than the earth's atmosphere. These TOA reflectance values include contributions from clouds and atmospheric aerosols and gases.

Following atmospheric corrections, seven bands of Landsat 8 OLI for each image were extracted as a layer stacking of multispectral bands image. The selection of ground control points (GCP) to be easy, each corrected image was subset to much small area. As images underwent unsupervised classification process, pixels were grouped based on the reflectance properties which are known as “clusters”. Each cluster with different land cover classes were identified and multiple clusters have been merged into a land cover type in order to yield subset and mask out the wetted bodies from satellite data as illustrated in Figure-4.

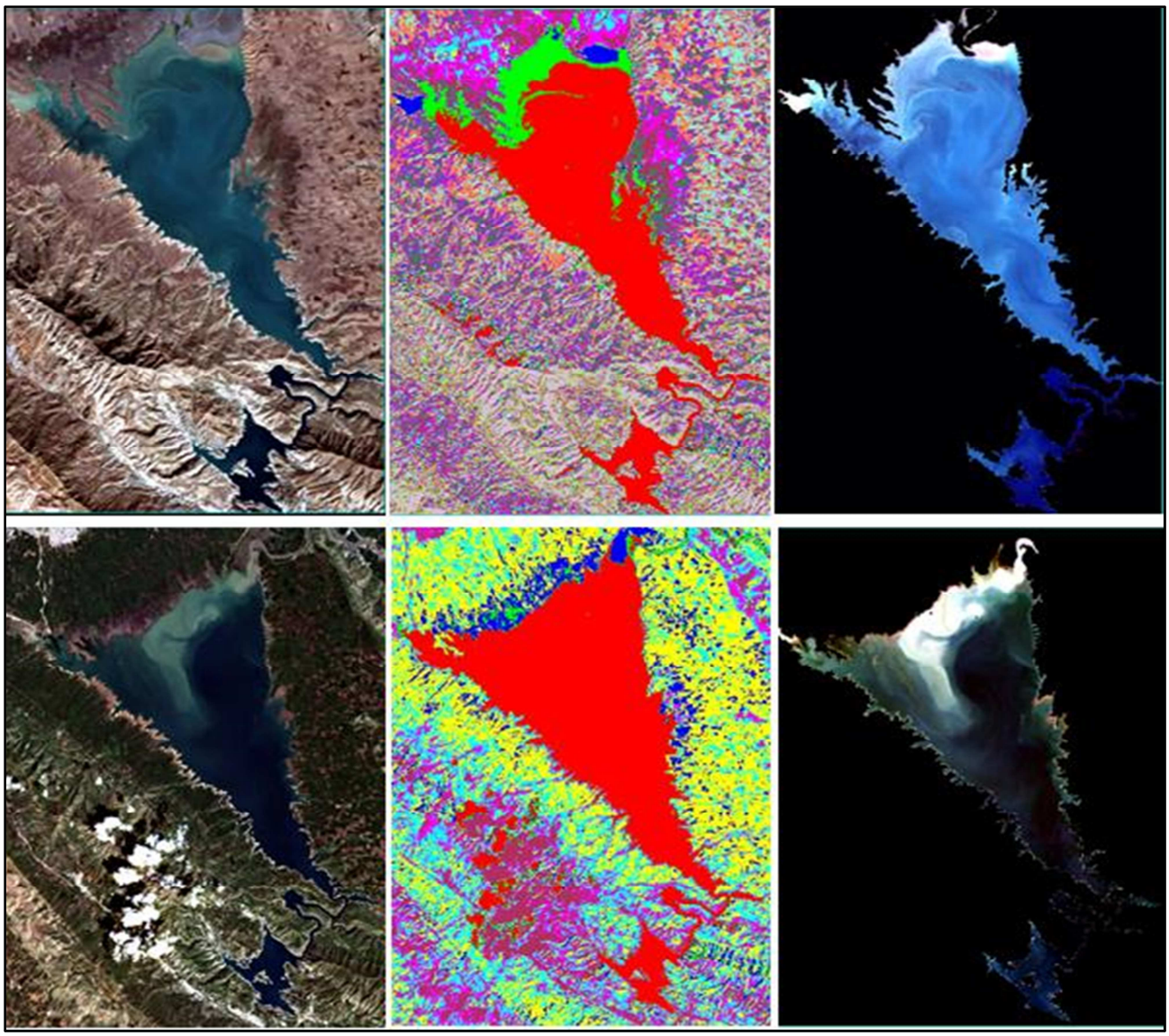

Figure-4: Landsat 8 OLI image on October 24th, 2014 (upper) and April 2nd, 2015 (lower), after subset (left), unsupervised classification (middle), after extraction of wetted area (right).

To ensure the accuracy of the classification, an accuracy assessment of the land use / land cover map was carried out to compare certain pixels in classified subsets with reference pixels through using supervised classification. The supervised classification was determined by integrated approach based on technique used by ENVI software to improve the accuracy of predictive models.

Finally, transformation process to the satellite image (extracted studied area) was provided which typically involves the manipulation of multiple bands of data. Before transformation, image enhancement methods are applied separately to each band of a multi-spectral image such as contrast enhancement, linear contrast stretch, and spatial filtering.

Image transformation in this study creates an output dataset where each output band is a linear combination of all the input bands. The first transform is Independent Component Analysis method (ICA) which works well with hyperspectral data because it is more likely to treat sparse targets as important features. However, ICA method can take a significantly longer time to process. The next method is PCA which creates a number of PC bands, which are linear combinations of the original spectral bands that are uncorrelated. It can be calculated the same number of output PC bands as input spectral bands. The first PC band contains the largest percentage of data variance and the second PC band contains the second largest data variance, and so on. Final transformation method is MNF which is a linear transformation uses separate PCA rotations to segregate noise in the data and to reduce the dimensionality of the original dataset.

E. Band Ratios

Band Ratios which enhance the spectral differences between bands were used in the present study to increase and provide more independent variables for the models and to reduce the effects of topography. Dividing one spectral band by another produces an image that provides relative band intensities and enhances the spectral differences between bands. One of the band that has been used is Coastal Blues which is Landsat 8 OLI's new band that not available in Landsat 7. It was used with other spectral bands whether as a combination or band ratios such as (C/B), (C/G), (C/R), (C/NIR), (B/C), (B/G), (B/R), (B/NIR), (G/C), (G/B), (G/R), (G/NIR), (R/C), (R/G), (R,B), (R/NIR), (NIR/C), (NIR/B), (NIR/G), (NIR/R), (B+C), (B+G), (B+R), (B+NIR), (G+C), (G+R), (G+NIR), (R+C), (R+NIR), (C+NIR), (B+G+C), (B+G+R), (B+G+NIR), (B+R+NIR), (B+R+C), (B+C+NIR), (G+C+R), (G+C+NIR), (R+C+NIR), (G+R+NIR), (B+G+R+NIR), (C+B+G+R), (C+B+G+NIR), (C+G+R+NIR), (C+B+R+NIR), (C+B+G+R+NIR), (B/NIR+C), (B/NIR+B), (B/NIR+G), (B/NIR+R), (B/NIR+NIR), (B/R+C), (B/R+B), (B/R+G), (B/R+R), (B/R+NIR), (NIR/B+C), (NIR/B+B), (NIR/B+G), (NIR/B+R) and (NIR/B+NIR). Then the spectral indices are combinations of surface reflectance at two or more wavelengths that indicate relative abundance of features of interest. Vegetation index (NDVI) is the most popular type, but other indices such as (LSWI), (MNDWI), (MSI), (NBR), (DVI), (IPVI), (NDMI) and (RVI) are available for water features which are used in this study.

F. Statistical Process

The statistical process focuses on the nature of statistical relationship between water quality and spatial response of satellite image for predicting the parameters. This process includes two parts, the first part is correlations, which examines linear relationships founds in the data that used to construct a scatterplot in a symmetric manner. While the second part, regression, considers the relationship of a response variable as determined by one or more explanatory variables.

A statistic that quantifies the strength of the linear relationship between the two variables is the correlation coefficient. The Pearson product-moment correlation coefficient is more widely used in measuring the association between two variables [19], and it has been taken into account in this study as well. There are three types of regression, simple, multiple and nonlinear regressions that can be used to determine the relationship between two or more variables [20]. The simple and multiple regressions were used in this study.

Many criteria for choice of subset size have been proposed and based on the principle of parsimony which suggests selecting a model with small residual sum of squares with as few parameters as possible [21]. The commonly used criteria are coefficient of determination (R^2), adjusted R^2 (R^2_{adj}), residual mean square MS_{res}, Mallows C_p Statistic, Akaike Information Criterion (AIC), Schwarz Bayesian Criterion (SBC), …, etc. [22]. In order to evaluate the performances of the developed models and to provide an indication of goodness of fit between the observed and calculated values, coefficient of determination (R^2), were used for each model investigated in this study.

The atmospherically corrected water masked Landsat 8 OLI images were used for water quality spatial analysis in this study. Twenty measurements points were located on the image based on the UTM coordinates determined with a GPS during water sampling and extracted the spectral band of used image that

mentioned before as independent variables, in ArcGIS Desktop software, to be used in correlation and regression modeling. Only the single pixel value for each sampling station was extracted.

A Pearson correlation matrix was first developed to determine the strengths of correlation between water quality parameters and spectral bands. Multiple linear regression methods were subsequently used to further explore the correlation between water quality parameters and spectral bands. The spectral band values at each measurement station were extracted from images for use as independent variables in bivariate regression models. Only the first seven bands that known as OLI were used for the analysis based on examination of the data. Independent variables include 14 surface reflectance bands, 7 PCA bands, 7 ICA bands,7 MNF bands, 20 band ratios, 26 band combinations and finally 9 indices. Dependent variables that were analyzed include some water quality parameters. All statistical analysis in this study has been done using SPSS computer software. Regression models were chosen to estimate the water quality parameters in Dokan Lake based on correlation coefficients and coefficients of determinations. Maps have been produced for measured versus computed water quality parameters values and visual interpretation of parameters.

Experimental Works

Theoretical study of any phenomenon, often, requires verification of its results empirically based on experimental measurement, field and laboratory, results. The experimental works in this study include collecting samples, testing and analyzing; the process which covers calculating the water quality parameters concentrations as described later. Before collecting the samples, first, twenty sampling stations were established on the map of Dokan lake. Twenty sampling stations were selected in the study area, with slightly equal distance from each other, in order to sufficiently estimate the quality of a water body of this size. Second, a GPS receiver (Garmin 62S) was used in a boat to locate each sample station in the lake for collecting the water samples. Forty samples were taken from the predetermined stations at two different dates. The stations made up the ground reference data of water quality parameters that were collected in a day which was coincident with the acquisition date of Landsat 8 which overpass and capture image. A small difference between the sampling and image capturing time has been recorded and this does not affect the accuracy of the analyses results because the concentration change of water quality parameters in several hours is very rare.

Water samples were collected on 24 October of 2014 as an autumn season and 02 April 2015 as a spring season. For autumn season, the sampling day was sunny with a temperature in the range 13-22 C^0 and a slight wind with a speed of 2.1 m/s coming from 159^0 N. On the sampling day, inflow discharge was 70 m^3/s same as the rate of outflow discharge which was less than the average daily because there was no precipitation in the catchment area. While for spring season, the sampling day was partly overcast but there is no cloud at the same area of Dokan lake with a temperature in the range 9-19 C^0 and a slight wind coming with the speed of 1.6 m/s from the146 N^0. The inflow discharge to the lake was 285 m^3/s which is higher than rate of outflow discharge and less than the average daily because there is no precipitation in the catchment area. The stations of sampling on October 24th, 2014 and April 2nd, 2015 in Dokan lake are shown in Figure-5.

At each station, two bottles of 250 ml samples (one was dark and the other was transparent) were taken for DO and BOD test (Winkler method). A bottle of 200 ml was taken for determining of E-choli test and two plastic bottles, which were approximately 500 ml of water sample, were collected for other parameters tests at 20 to 25 cm below the water surface where the temperature and sunlight may have no effect on them. To prevent the deterioration of algae and other organic matter, the sample bottles were kept in an ice cooler until analyzed in the laboratory for E-coli and BOD. The collected water samples from the 20 point stations within Dokan lake, in both dates, were tested and analyzed in three laboratories (except water temperature and pH which were measured in the field), to specify some of the physical, chemical and biological characteristics for the water of these samples. These tests were achieved by chemical process or by using recent portable instruments for water quality tests. The measured water quality parameters are: water temperature (T), pH, Turbidity, total suspend solid (TSS), total dissolved solid (TDS), electrical conductivity

(EC), nitrate (NO_3), phosphate (PO_4), dissolved oxygen (DO), biochemical oxygen demand (BOD) and biological characteristics which indicate the presence of E. coli in the water samples. Test results for water quality parameters correspond to the stations numbers on October 24th, 2014 and on April 2nd, 2015 are shown in Table-2 and Table-3.

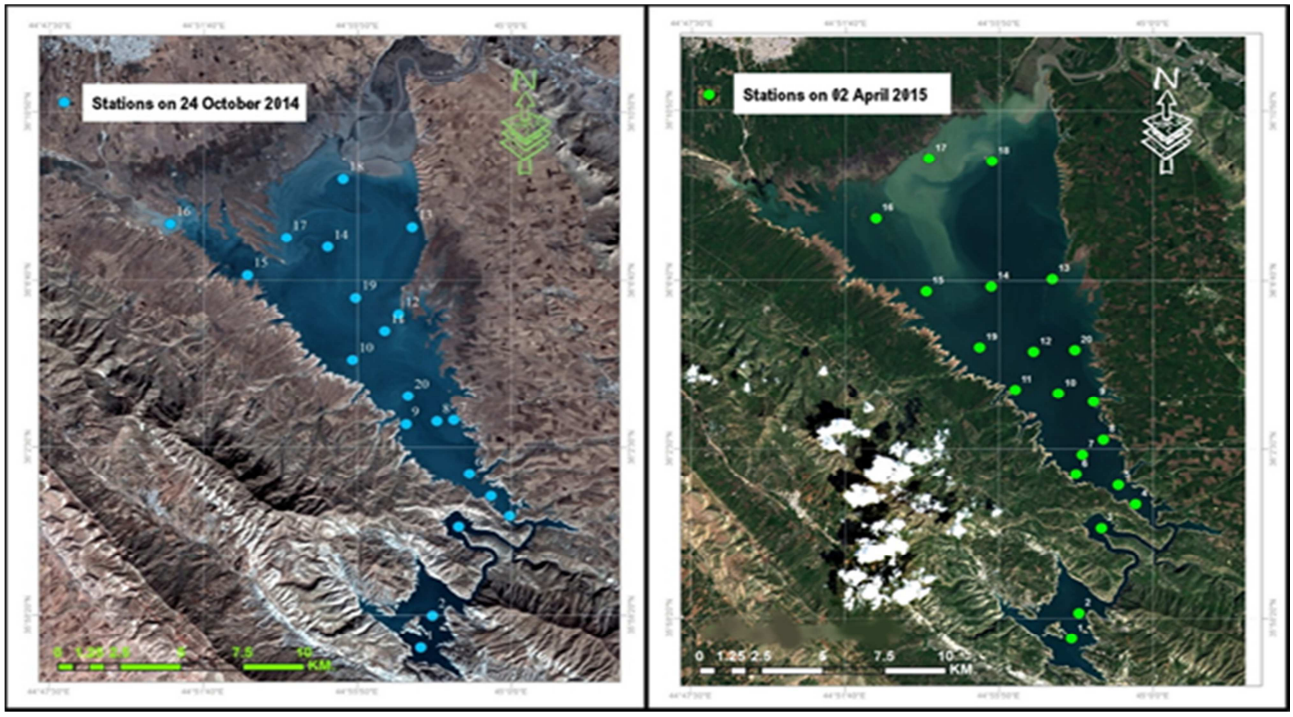

Figure-5: Station points of sampling on October 24th, 2014 and April 2nd, 2015 in Dokan lake.

Table-2: Test results for water quality parameters correspond to the stations numbers on October 24th, 2014.

ID	pH	NO_3 (mg/l)	EC (µhos/cm)	TDS (mg/l)	Temperature (C°)	Turbidity (NTU)	DO (mg/l)	PO_4 (mg/l)	TSS (mg/l)	M.P.N of E. coli	BOD (mg/l)	NO_3-N (mg/l)	TP (mg/l)
1	8.6	5.2	523	261.5	22.0	1.49	8.8	0.22	1.788	+ve	2.8	1.17	0.07
2	8.0	4.8	350	175	21.5	1.42	8.0	0.2	1.704	+ve	1.2	1.09	0.07
3	8.0	4.7	353	176.5	20.0	1.42	7.6	0.17	1.704	+ve	0.8	1.06	0.06
4	8.0	4.9	346	173	21.0	1.7	7.4	0.17	2.04	+ve	2.4	1.11	0.06
5	8.0	5.1	335	167.5	21.0	1.6	6.8	0.18	1.92	+ve	1.2	1.15	0.06
6	8.1	5.2	346	173	20.5	2.2	6.4	0.18	2.64	+ve	1.2	1.17	0.06
7	8.1	4.5	350	175	21.0	3	6.8	0.20	3.6	+ve	2.4	1.02	0.07
8	8.1	4.7	112.5	56.25	21.0	2.9	7.0	0.22	3.48	+ve	3.6	1.06	0.07
9	8.0	5.3	343	171.5	20.5	3.01	6	0.20	3.612	+ve	2.0	1.20	0.07
10	8.2	5.1	346	173	20.5	2.8	6.8	0.20	3.36	+ve	1.6	1.15	0.07
11	8.1	3.5	319	159.5	20.5	3	5.6	0.20	3.6	+ve	2.8	0.79	0.07
12	8.0	5.2	353	176.5	20.5	3.3	5.4	0.22	3.96	+ve	0.4	1.17	0.07
13	8.2	3.8	344	172	21.0	2.8	6.0	0.22	3.36	+ve	2.0	0.86	0.07
14	8.2	4.5	333	166.5	20.5	3.1	5.6	0.26	3.72	+ve	0.8	1.02	0.08
15	8.3	5.4	351	175.5	20.5	5	7.0	0.32	6.00	+ve	0.4	1.22	0.10
16	8.2	1.3	455	227.5	19.0	58	7.6	0.30	69.6	+ve	1.2	0.29	0.10
17	8.4	0.5	347	173.5	20.5	8.7	6.8	0.24	10.44	+ve	0.4	0.11	0.08
18	8.2	1.6	353	176.5	21.0	3.37	6.6	0.29	4.04	+ve	2.0	0.36	0.09
19	8.1	1.7	332	166	20.5	3.8	5.4	0.27	4.56	+ve	1.2	0.38	0.09
20	8.1	1.0	335	167.5	21.0	3.8	7.4	0.22	4.56	+ve	1.2	0.23	0.07

Table-3: Test results for water quality parameters correspond to the stations numbers on April 2nd, 2015.

ID	pH	NO_3 (mg/l)	EC (µhos/cm)	TDS (mg/l)	Temperature (C°)	Turbidity (NTU)	DO (mg/l)	PO_4 (mg/l)	TSS (mg/l)	M.P.N of E. coli	BOD (mg/l)	NO_3-N (mg/l)	TP (mg/l)
1	8.7	3.286	413	206.5	14.0	2.60	2.4	0.244	3.12	+ve	0.1	0.74	0.08
2	8.6	3.413	387	193.5	14.0	3.90	2.4	0.244	4.68	+ve	0.1	0.77	0.08
3	8.5	3.249	391	195.5	14.0	2.01	2.5	0.243	2.412	+ve	0.2	0.73	0.08
4	8.6	3.522	365	182.5	14.5	2.65	2.3	0.248	3.18	+ve	0.3	0.80	0.08
5	8.6	0.156	380	190.0	14.0	1.25	2.6	0.52	1.5	+ve	0.5	0.04	0.17
6	8.6	3.322	421	210.5	14.0	2.65	2.7	0.287	3.18	+ve	0.5	0.75	0.09
7	8.6	3.177	384	192.0	14.0	2.20	2.4	0.27	2.64	+ve	0.3	0.72	0.09
8	8.6	3.05	374	187.0	14.0	1.80	2.6	0.239	2.16	+ve	0.3	0.69	0.08
9	8.6	3.268	389	194.5	14.5	3.20	2.8	0.41	3.84	+ve	0.6	0.74	0.13
10	8.6	3.286	403	201.5	14.5	2.20	2.7	0.423	2.64	+ve	0.6	0.74	0.14
11	8.7	3.140	412	206.0	15.0	1.90	2.3	0.331	2.28	+ve	0.2	0.71	0.11
12	8.7	3.340	402	201.0	14.0	1.40	2.8	0.253	1.68	+ve	0.2	0.75	0.08
13	8.6	3.249	383	191.5	14.5	1.16	2.4	0.253	1.392	+ve	0.3	0.73	0.08
14	8.7	3.504	397	198.5	14.5	1.40	2.3	0.27	1.68	+ve	0.1	0.79	0.09
15	8.5	2.142	430	215.0	15.0	1.70	2.2	0.26	2.04	+ve	0.2	0.48	0.08
16	8.5	1.307	416	208.0	16.0	2.60	2.5	0.25	3.12	+ve	0.5	0.30	0.08
17	8.5	1.833	390	190.0	16.0	2.11	2.3	0.244	2.532	+ve	0.8	0.41	0.08
18	8.6	3.122	389	189.5	17.0	1.50	2.6	0.218	1.8	+ve	0.4	0.70	0.07
19	8.6	3.013	355	177.5	15.0	1.80	2.3	0.266	2.16	+ve	0.2	0.68	0.09
20	8.6	3.322	428	214.0	15.0	1.16	2.9	0.211	1.392	+ve	0.5	0.75	0.07

Results and Discussions

Satellite image data undergoes several processing and transformation, then, by using statistical correlation and developing models, water quality parameters are estimated. The processes depend on the spectral reflectance characteristics of the water in OLI bands of the Landsat 8 spectral bands. In the present study, the values of water quality parameters obtained from the developed models were studied and compared to the measured field and laboratory data. Twelve parameters were selected as dependent variables at twenty station points to generate cartographic maps and water quality estimation for two different seasons (autumn and spring) for the entire area of Dokan lake. The surface reflectance of Landsat 8 OLI bands for selected stations in Dokan lake on October 24th, 2014 and April 2nd, 2015 are shown in Figure-6 and Figure-7.

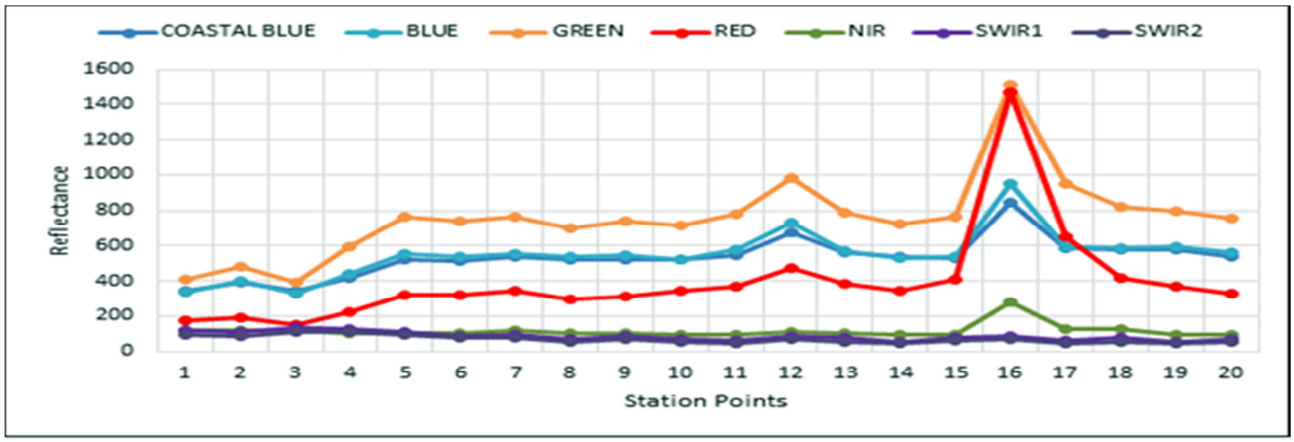

Figure-6: Surface reflectance of Landsat 8 OLI bands for selected stations in Dokan lake on October 24th, 2014 (autumn season).

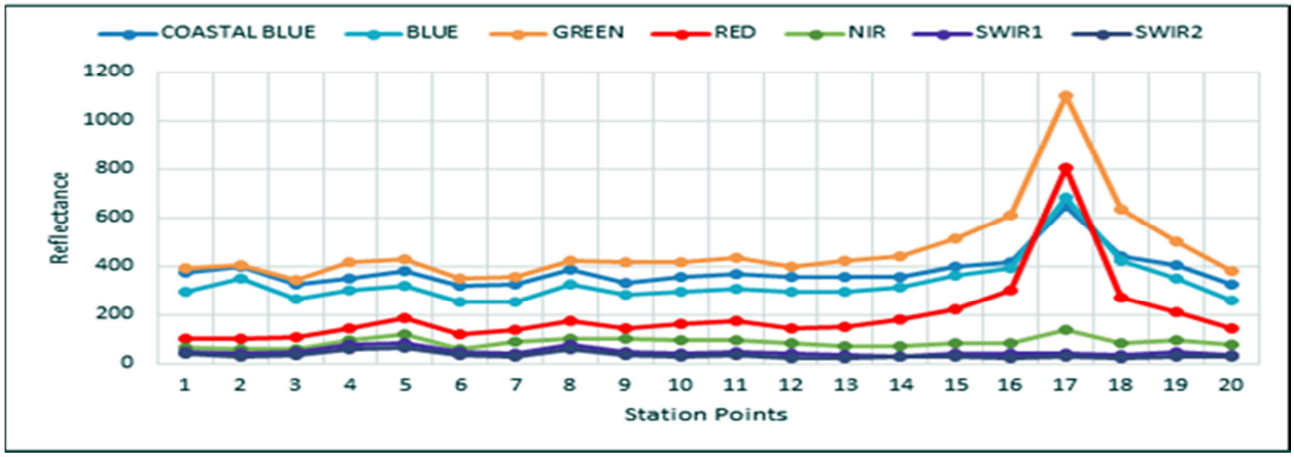

Figure-7: Surface reflectance of Landsat 8 OLI bands for selected stations in Dokan lake on April 2nd, 2015 (spring season).

After the image processing, the multiple linear regressions were used to establish the relationship between the water quality parameters as dependent variables and Landsat 8 OLI spectral data as independent variables. Spectral data of autumn and spring images and in situ measured data on October 24th, 2014 and April 2nd, 2015 of water quality parameters correlated and multiple linear regression models were developed and selected based on the R^2 for estimating the pH, EC, TDS, T, NO_3, $NO_3 - N$, PO_4, TP, TSS, Turbidity, DO and BOD water quality parameters as shown in Table-4.

Table-4: WQPs models for autumn season on October 24th, 2014 and spring season on April 2nd, 2015 for Dokan lake.

Season	Water Quality Parameters Models	R^2
Autumn	$pH = 10.491 - 0.051\ (ICA\ SWIR1) - 0.742\ (G/R) - 0.147\ (G: Float)$	0.80
	$EC = 241.50 + 529.504\ (NIR/C)$	0.70
	$T = 21.765 - 0.001\ (C + R)$	0.73
	$NO_3 = 4.678 - 10.400\ (NBR) + 0.004\ (B/NIR + B)$	0.81
	$NO_3 - N = 0.772 - 2.445\ (LSWI) + 0.001\ (B/NIR + B)$	0.82
	$PO4 = 0.401 - 0.083\ (G/R)$	0.70
	$TP = 0.127 - 0.025\ (G/R)$	0.75
	$TSS = -50.077 + 0.063\ (R + NIR) + 16.708\ (C/R)$	0.99
	$NTU = -42.564 + 0.052\ (R + NIR) + 13.923\ (C/R)$	0.99
	$DO = 10.841 - 0.682\ (C/NIR) - 0.002\ (B/NIR + B)$	0.83
	$TDS = 120.75 + 264.752\ (NIR/C)$	0.70
	BOD(No Model)	-
Spring	$pH = 8.485 + 0.045\ (ICA\ C)$	0.28
	$EC = 422.034 - 1080.365\ (SWIR1: Float)$	0.21
	$T = 10.516 + 0.386\ (ICA\ G) + 2.933\ (G/B)$	0.86
	$NO_3 = 4.043 - 2.068\ (R/C)$	0.55
	$NO_3 - N = 0.913 - 0.469\ (R/C)$	0.56
	$PO_4 = 0.146 + 0.552\ (NIR/C)$	0.21
	$TP = 0.089 + 0.016\ (ICA\ NIR)$	0.32
	$TSS = -5.068 + 1.695\ (B/R) + 3.548\ (G/C)$	0.46
	$NTU = -4.223 + 1.412\ (B/R) + 2.957\ (G/C)$	0.46
	DO (No Model)	-
	TDS (No Model)	-
	$BOD = -1.037 + 1.062\ (G/B) + 0.003\ (PCA\ B)$	0.57

For spring season, the R^2 value for some water quality parameters models were decreased due to decreasing the reflectance of Landsat 8 OLI that captured on April 2nd, 2015 and seasonal changes of water quality parameters of Dokan lake. The reflectance reduction is due to the light penetration from water

surface, intensity of incipient of light, angle of ray incidence and scattering and absorption light within the water. The penetration of light anywhere in the water is reduced either when the sun is away from zenith, that means sun elevation for October 24th, 2014 is 52.958263^0 is more than the sun elevation of April 2nd, 2015 which is 39.77423027^0. In the study area, all the water quality parameters showed a wide variation in space and time. Temporal variations were due to seasonal influences mainly, the effect of rainfall. The data analyses revealed that most of the parameters showed a substantial decrease in correlation with the models after winter season.

In autumn season, the best pH model shows that the pH value is correlated with (ICA SWIR1), (G/R) and (G: Float) and has highest R^2 value of 0.80 while in the spring season pH models no longer correlated with the spectral band due to higher variability of the values. The best model that selected has R^2 value of 0.28 which is use (ICA C) as independent variable.

The total dissolved solids (TDS) values for autumn season are highly correlated with band-ratios of NIR to coastal blue band (NIR/C) of Landsat 8 OLI. The TDS values in autumn season include an outlier value in station point 8. On the other hand, for spring season there is no model obtained because of the change in seasonal variations and decreasing in conductivity, while they affected by average temperatures, they also affected by water flow which increases in this season.

In this study, the temperature (T) have been developed for estimating and predicting its value. In autumn season, the temperature has been correlated with combination of coastal blue and red reflectance band of Landsat 8 OLI and a multiple linear regression model has been selected with R^2 value of 0.73 for estimating water temperature values in Dokan lake. For spring season, another multiple linear regression model has developed and the highest R^2 of 0.86 is obtained with (ICA G) and (PCA B) as independent variables.

In the present study, nitrogen compounds of water are represented in two forms, nitrate (NO_3) and nitrate-nitrogen ($NO_3 - N$). For NO_3 the best model with highest R^2 value of 0.81 was obtained when the values for autumn season are correlated with (NBR) index and (B/NIR + B) of Landsat 8 OLI. In addition, for autumn season the best $NO_3 - N$ model with highest R^2 value of 0.82 was obtained when the values are correlated with (LSWI) index and (B/NIR + B). On the other hand, for spring season, each one of NO_3 and $NO_3 - N$ values are correlated with (R/C) with R^2 values of 0.55 and 0.56 respectively. The computed values of NO_3 and NO_3 – N were high at the inlet of the lake and decreased towards Dokan Dam. However, in general, the values decreased in spring season due to seasonal factors such as dilution due to rainfall, aeration, etc. Also, phosphate represented in two forms as phosphate (PO_4) and total phosphate (TP). The best models for PO_4 and TP with highest R^2 values of 0.70 and 0.75 respectively were obtained for autumn season when PO_4 and TP values were correlated with band-ratio (G/R) of Landsat 8 OLI. However, for spring season, the models of PO_4 and TP are weakly correlated with spectral bands of Landsat 8 OLI. The computed PO_4 and TP were decreased from inlet of the lake towards Dokan Dam. Generally, spring season values decreased due to seasonal factors such as dilution due to rainfall, aeration, etc.

In the present study, Dissolved Oxygen (DO) is represented as concentration (mg/l) and the best model for autumn season has R^2 value of 0.83 when DO values are correlated with band-ratios (C/NIR) and (B/NIR + B). On the other hand, for spring season no DO models were obtained because of the changes such as aeration and plant growth in the lake.

Furthermore, BOD models only developed for spring season data and the best model obtained by using the band ratio (G/B) and (PCA B) with highest R^2 value of 0.57. In autumn season, DO concentration values that estimated from the model ranged from 9.45-10.84 mg/l at the lake and decreased to less than 3.6 mg/l at center of the lake. This might be due to the diffusion and aeration, photosynthesis, respiration and decomposition.

For autumn season data, the best models for *TSS* and Turbidity were obtained by correlating total suspended solid and turbidity values with (R+NIR) and band-ratio of (C/R) of Landsat 8 OLI with a highest R^2 value of 0.99. However, for spring season, the best models have R^2 value of 0.46 which obtained by correlating total suspended solid and turbidity values with band-ratios of (B/R) and (G/C). The performance

reduction of the model in this season is due to the outlying value of TSS (high value) at station point 16 which is close to lake margin. The TSS and Turbidity values in autumn season include an outlier value in station point 16, therefore, the value of this point was assumed as a missed value in developing the models in order to increase the R^2 value. The estimated values of turbidity are decreased from the lake inlet toward the Dokan dam. The higher values at the inlet are due to soil erosion, runoff, discharges and stirred bottom sediments or algal blooms.

For EC in Dokan lake, the best model with highest R^2 value of 0.70 shows that the EC values for autumn season is only correlated with band-ratios of (NIR/C) band of Landsat 8 OLI which is a new band that is not available in Landsat 7 ETM. However, for spring season model the R^2 value is 0.21 with the band-ratio of (NIR/C) as independent variable and this may be because of seasonal change of EC values and change in solar and reflectance data of satellite image. The EC values in autumn season include an outlier value in station point 8, therefore, the value of this point was assumed as a missed value in developing the models in order to increase the R^2 value.

The measured and computed values from the best models for water quality parameters and the maps of generated values in Dokan lake for autumn and spring seasons are shown in Figure-8 through Figure-10.

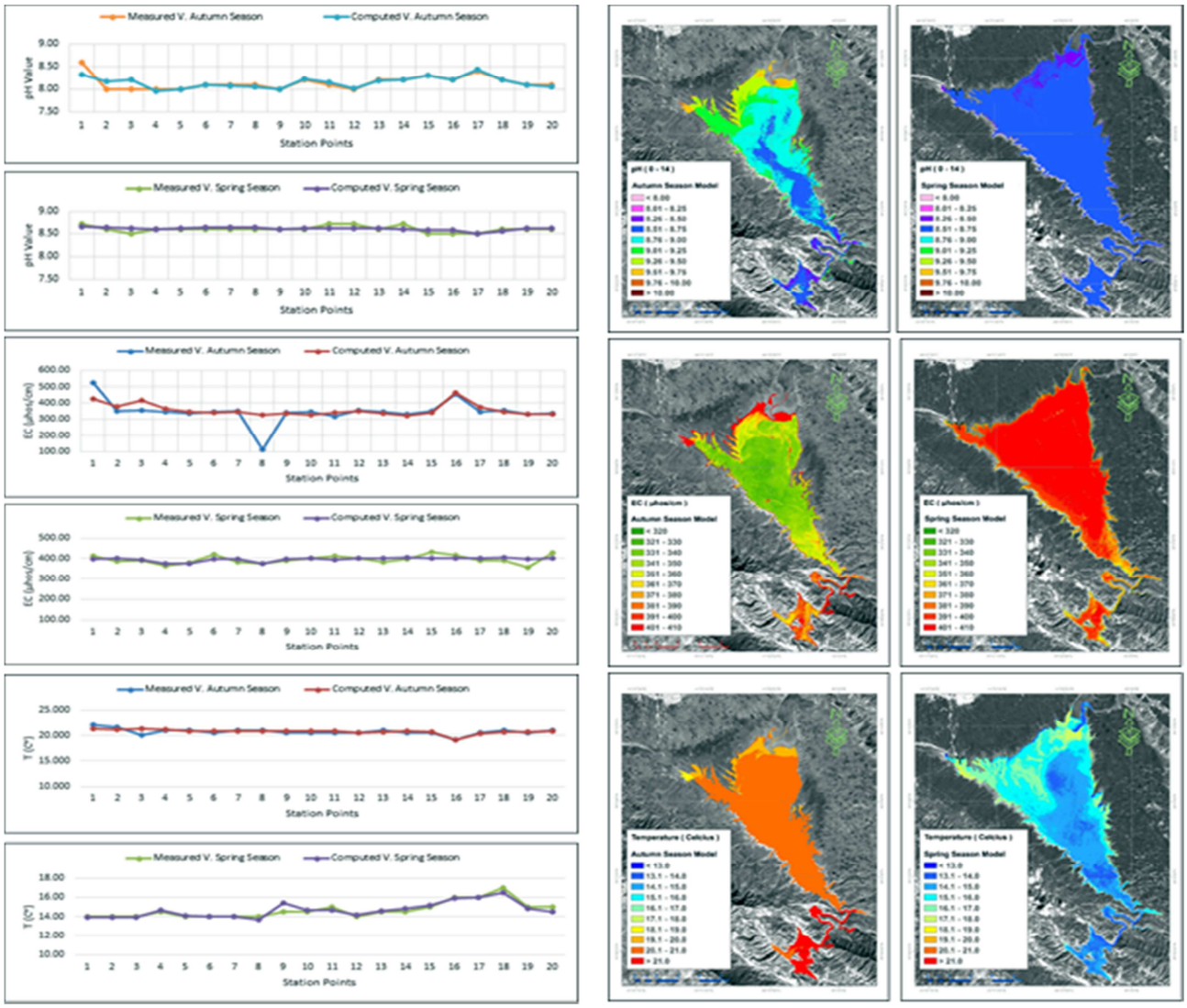

Figure-8: Measured and computed values (Left) and mapped computed values (Right) of pH, EC and T in Dokan lake for autumn and spring seasons.

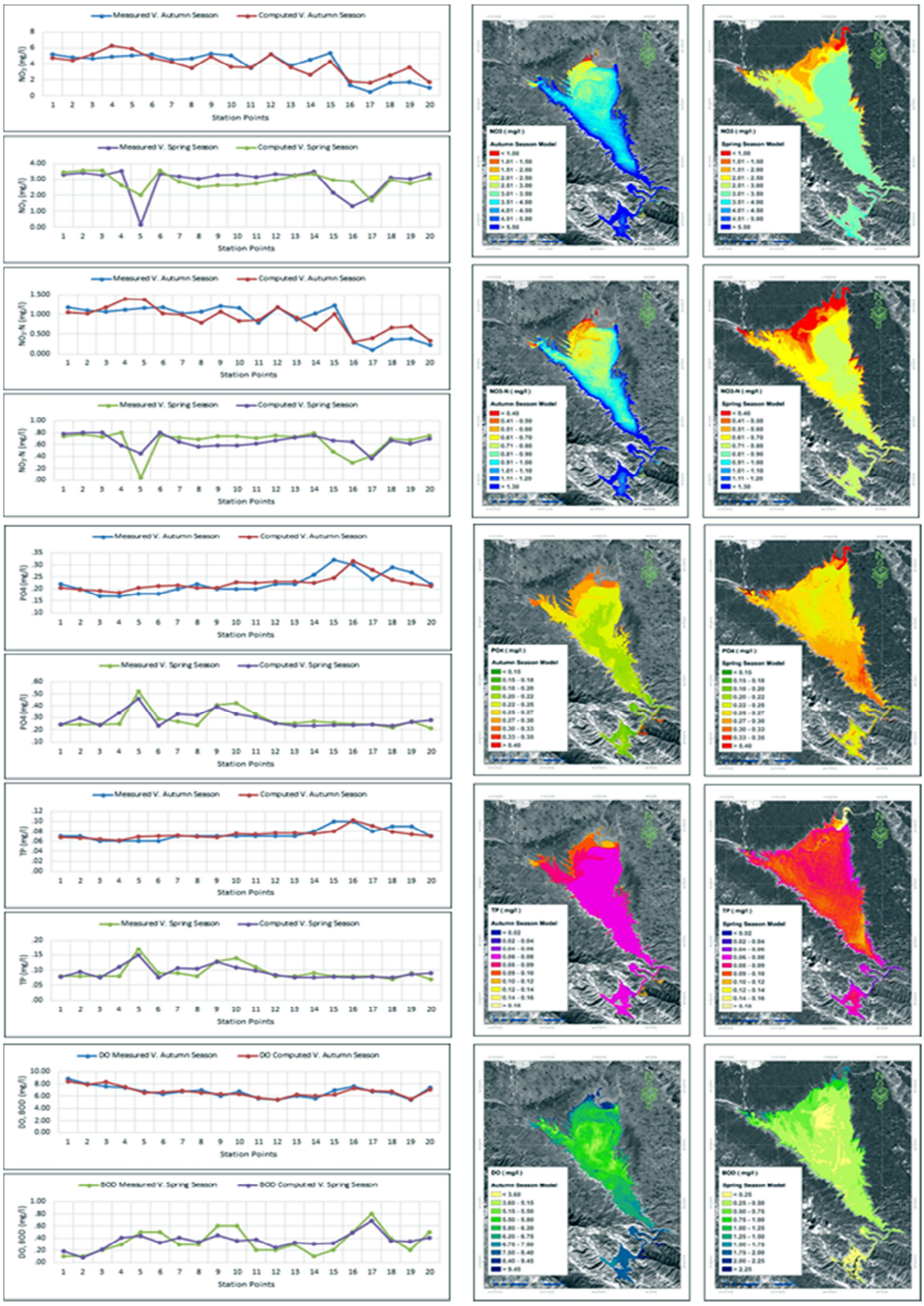

Figure-9: Measured and computed values (Left) and mapped computed values (Right) of NO_3, NO_3-N, PO_4, TP, DO and BOD in Dokan lake for autumn and spring seasons.

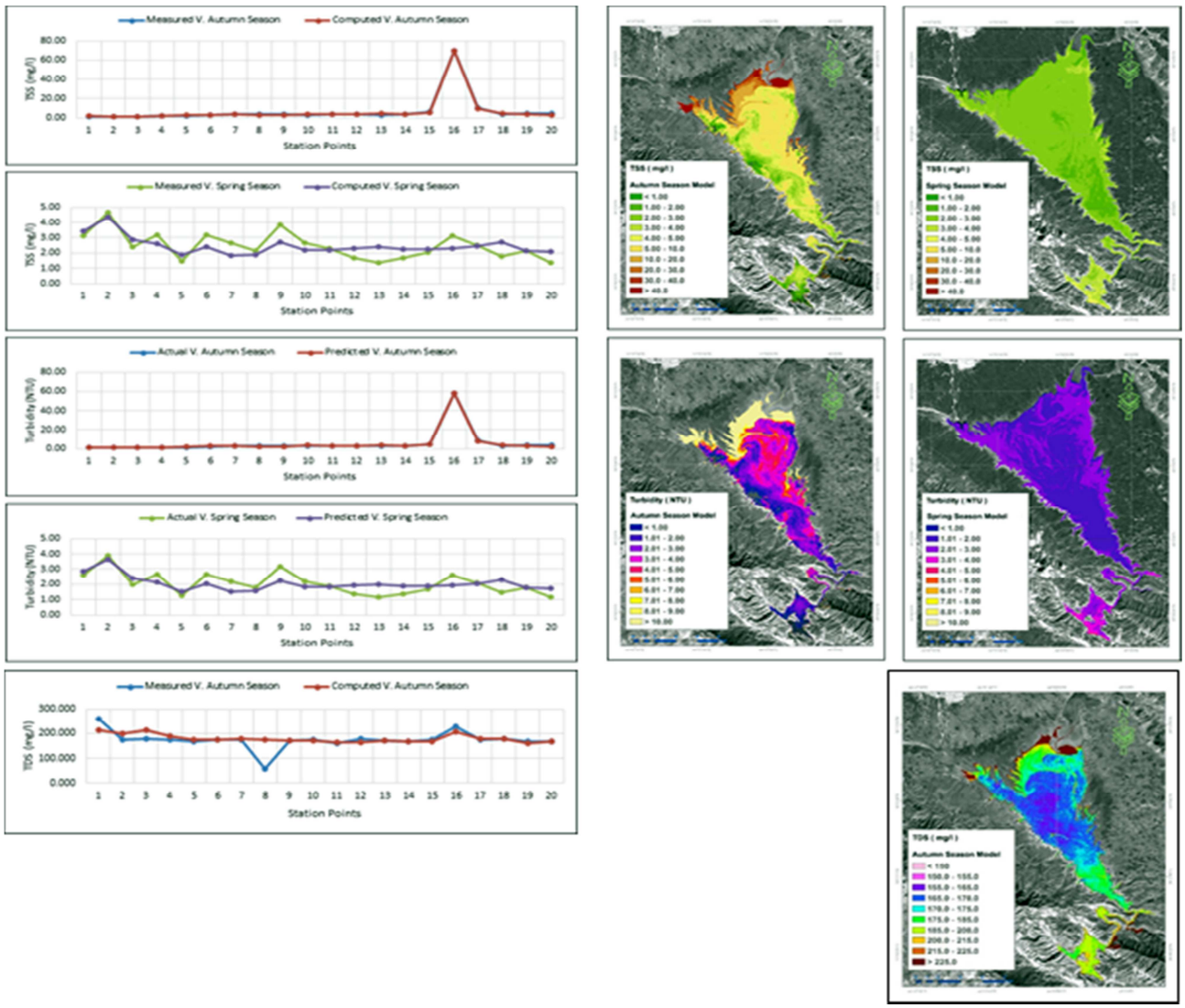

Figure-10: Measured and computed values (Left) and mapped computed values (Right) of TSS, Turbidity and TDS in Dokan lake for autumn and spring seasons.

To check the generality of the developed models, the models that developed for autumn season data have been tested by using the spring season data and vice versa. The best models that have been developed for autumn season data on October 24th, 2014 were used to predict the water quality parameters of Dokan lake water for spring season data on April 2nd, 2015. A new R^2 has been calculated for each model as shown in Table-5. On the other hand, the best models that have been developed for spring season data on April 2nd, 2015 were used to predict the water quality parameters of Dokan lake water for autumn season on October 24th, 2014. Also, a new R^2 has been calculated for each model as shown in Table-5. The results show that the values of R^2 are decreased to less than or equal 0.5. The main factor which causes to change the performance of models is high variation of all data when the season is changed. As clear from the results, no model can be used as general model for all seasons of the year.

Table-5: Generality of autumn and spring seasons models of water quality parameters for Dokan Lake (autumn models used for spring data and vice versa).

Season	Water Quality Parameters Models	R^2
Autumn	$pH = 10.491 - 0.051(\text{ICA SWIR1}) - 0.742(G/R) - 0.147(\text{G: Float})$	0.47
	$EC = 241.500 + 529.504\,(\text{NIR}/C)$	0.42
	$TDS = 120.750 + 264.752\,(\text{NIR}/C)$	0.42
	$T = 21.765 - 0.001\,(C + R)$	0.50
	$NO_3 = 4.678 - 10.400\,(\text{NBR}) + 0.004\,(B/\text{NIR} + B)$	0.46
	$NO_3 - N = 0.772 - 2.445\,(\text{LSWI}) + 0.001\,(B/\text{NIR} + B)$	0.45
	$PO_4 = 0.401 - 0.083\,(G/R)$	0.41
	$TP = 0.127 - 0.025\,(G/R)$	0.41
	$TSS = -50.077 + 0.063\,(R + \text{NIR}) + 16.708\,(C/R)$	0.50
	$NTU = -42.564 + 0.052\,(R + \text{NIR}) + 13.923\,(C/R)$	0.50
	$DO = 10.841 - 0.682\,(C/\text{NIR}) - 0.002\,(B/\text{NIR} + B)$	0.50
	BOD (No Model)	-
Spring	$pH = 8.485 + 0.045\,(\text{ICA C})$	0.46
	$EC = 422.034 - 1080.365\,(\text{SWIR1 : Float})$	0.42
	TDS (No Model)	-
	$T = 10.516 + 0.386\,(\text{ICA B3}) + 2.933\,(G/B)$	0.50
	$NO_3 = 4.043 - 2.068\,(R/C)$	0.36
	$NO_3 - N = 0.913 - 0.469\,(R/C)$	0.36
	$PO_4 = 0.146 + 0.552\,(\text{NIR}/C)$	0.37
	$TP = 0.089 + 0.016\,(\text{ICA NIR})$	0.42
	$TSS = -5.068 + 1.695\,(B/R) + 3.548\,(G/C)$	0.08
	$NTU = -4.223 + 1.412\,(B/R) + 2.957\,(G/C)$	0.08
	DO (No Model)	-
	$BOD = -1.037 + 1.062\,(G/B) + 0.003\,(\text{PCA B})$	0.39

Conclusions

From the results analyses of experimental works, remote sensing processing works and WQPs models in the present study, one can observe and conclude that all bands of Landsat 8 OLI can be correlated with water quality parameters especially the new band (Coastal Blue) with its combinations and band-ratios, independent component analysis (ICA) and minimum noise fraction (MNF) which have not been used before.

The correlation of water quality parameters with reflectance are decreased in spring season due to seasonal changes (high variance between station points data). Regarding the temperature, its value not only correlated with radiance of satellite image data but also correlated indirectly with reflectance of satellite data and can be accurately computed and mapped in the lake because it has a correlation with other WQPs; due to the change in temperature cause the change in that parameters. The results show that no model can be used as general model for all seasons of the year.

References

[1] S. Y. Ming, J. M. Carolyn and M. S. Robert, *"Adaptive Short-Term Water Quality Forecasts Using Remote Sensing and GIS",* in AWRA Symposium on GIS and Water Resources, Ft. Lauderdale, FL, (1996).

[2] R. D. Hedger, P. M. Atkinson and T. J. Malthus, *"Optimizing Sampling Strategies for Estimating Mean Water Quality in Lakes Using Geostatistical Techniques with Remote Sensing*". Lakes & Reservoirs: Research and Management, Vol. 6, pp. 279-288. (2001).

[3] D. Doxaran, R. C. N. Cherkuru and S. J. Lavender, *"Use of Reflectance Band Ratios to Estimate Suspended and Dissolved Matter Concentrations in Estuarine Waters*". International Journal of Remote Sensing, Vol. 26, No. 8, pp. 1763-1769, (2005).

[4] E. Alparslan, C. Aydöner, V. Tufekci and H. Tüfekci, *"Water Quality Assessment At Ömerli Dam Using Remote Sensing Techniques*". Environ Monit Assess, Vol. 135, pp. 391-398. (2007).

[5] B. Nas, H. Karabork, S. Ekercin and A. Berktay, *"Assessing Water Quality in the Beysehir Lake (Turkey) by the Application of GIS, Geostatistics and Remote Sensing*", in Taal 2007: The 12th World Lake Conference, Konya, Turkey, (2008).

[6] P. Li, L. Jiang and Z. Feng, *"Cross-Comparison of Vegetation Indices Derived from Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and Landsat-8 Operational Land Imager (OLI) Sensors*". Remote Sensing, Vol. 6, pp. 310-329, (2014).

[7] F. D. v. d. Meer and S. M. d. Jong, *"Image Spectrometry ; Basic Principle and Prospective applications*", Third Edition ed., Donrecht, The Netherlands: Springer, (2006).

[8] K. Ararat, N. A. Hassan and S. Abdul Rahman, *"Key Biodiversity Survey of Kurdistan, Northern Iraq"*. Nature Iraq Report, Sulaimani, Kurdistan, Iraq, (2009).

[9] S. S. Ali and S. K. R. Salley, *"Numerical Groundwater Flow Modeling for the Intwrgranular Aquifer in Sarsian Sub-Basin, Dokan Lake, Iraqi Kurdistan Region*". Journal of Zankoy Sulaimani-Part A, Vol. 15, No. 1, pp. 125-141, (2003).

[10] A. H. A. Bilbas, *"Ecosystem Health Assessment of Dukan Lake, Sulaimani, Kurdistan Region of Iraq"*, Erbil: Salahaddin University, (2014).

[11] K. Navulur, *"Multispectral Image Analysis Using the Object-Oriented Paradigm*", First Edition, Boca Raton, FL: CRCpress, Taylor & Francis Group. LLC, (2007).

[12] T. M. Lillesand, R. W. Kiefeer and J. W. Chipman, *"Remote Sensing and Image Interpretation",* Six Edition ed., USA: John Wiley & Sons Inc., (2007).

[13] USGS, *"Landsat-A Global Land-Imaging Mission*", [Online]. Available: http://remotesensing.usgs.gov.. [Accessed 9 6 2015].

[14] UN-ESCWA and B. , *"Inventory of Shared Water Resources in Western Asia*", Beriut: United Nations Economic and Social Commission for Western, (2013).

[15] C. D. S. S. Baboo and S. Thirunavukkarasu, *"Geometric Correction in High Resolution Satellite Imagery Using Mathematical Methods: A Case Study in Kiliyar Sub Basin*". Global Journal of Computer Science and Technology: Graphics & Vision, Vol. 14, No. 1, pp. 35-40, (2014).

[16] S. C. Goslee, *"Analyzing Remote Sensing Data in R: The Landsat Package*". Journal of Statistical Software, Vol. 43, No. 4, pp. 1-25, (2011).

[17] C. C. f. R. S. CCRS, *"Tutorial: Fundamentals of Remote Sensing*". [Online]. Available: http://www.nrcan.gc.ca/earth-sciences/geomatics/satellite-imagery-air-photos/satellite-imagery-products/educational-resources/9309. [Accessed 03 07 2015].

[18] T. W. Ray, *"A FAQ on Vegetation in Remote Sensing"*. [Online]. Available: http://www.yale.edu/ceo/Documentation/rsvegfaq.html. [Accessed 25 12 2014].

[19] P. M. Berthouex and L. C. Brown, *"Statistics for Enviromental Engineers*", 2 ed., Boca Raton, London, New York, Washington, D.C.: Lewis Publishers, (2002).

[20] D. Campbell and S. Campbell, *"Introduction to Regression and Data Analysis*", StatLab Workshop Series, (2008).

[21] J. O. Rawlings, S. G. Pantula and D. A. Dickey, *"Applied Regression Analysis: A Research Tool*", Second Edition, Newyork: Springer, (1998).

[22] S. Weisberg, *"Applied Linear Regression"*, Third Edition ed., Hoboken, New Jersey: John Wiley & Sons, Inc., (2005).