Empirical estimation of surface water quality parameters in Lake Kerkini using Landsat ETM+/OLI

M. Kanj[1], I. Manakos[2*], I. Tsolakidis[3], N. Katselas[3], and C. Kalaitzidis[1]
[1] Department of Geoinformation in Environmental Management,
Mediterranean Agronomic Institute of Chania, 73100 Chania, Greece
[2] Information Technologies Institute, Centre for Research and Technology Hellas,
57001 Thessaloniki, Greece
[3] Lake Kerkini Management Authority, 62055 Kerkini, Serres, Greece

*Corresponding author: E-mail: imanakos@iti.gr, Tel: (+302311) 257 701-3, Fax: (+302310) 474 128

Abstract
Earth Observation data offer a unique opportunity to enhance capacity in open water surface resources monitoring from space. This needs to be quantitatively proven to achieve acceptability and integration to established decision-making processes. Lake Kerkini, in its function as flood regulator of Strymonas river, nesting site for migratory birds, irrigation water provider, feeding site for rare herbivores (e.g. *Bubalus bubalis*) and integrated in the protected areas network of Europe (Natura2000 & Ramsar Convention site), offers an advantageous testing bed for this activity. Landsat ETM+/OLI data were explored against in situ data to develop regression estimation models for temperature, Secchi disk depth, pH, dissolved oxygen, electric conductivity and nutrients. Results reached a coefficient of determination greater than 0.84 for some cases, even for employing the polynomial formulas across sensors (ETM or OLI), whereas for others, less accurate results are obtained (around 0.6 R^2), and in a few cases, failures are experienced. Transferability of results, challenges, and opportunities (e.g. application using up to 7 days distant image data) are presented and discussed.

Keywords: *water quality; regression analysis; Landsat; in-situ measurements; lake Kerkini.*

1. INTRODUCTION

Inland freshwater water bodies are important natural resources that support human well-being, in terms of providing food and shelter, climate and flood regulation, as well as providing income through tourism [1]. Any changes to these ecosystems, due to climate change, land degradation, unregulated withdrawal of surface and ground water, nutrient overload from agricultural or urban and industrial effluents, have led to the shortage and degradation of water quality [2–4]. The assessment and continuous monitoring of water quality in freshwater bodies can enhance understanding of the hydrochemical cycles and the effective management of water resources [4].

Water quality monitoring by conventional techniques depends on in-situ measurements or laboratory analysis of samples. These may provide accurate measurements, but they are usually time-consuming and costly, with the inability to investigate spatiotemporal variations and trend analysis of entire water bodies, especially in inaccessible locations. Monitoring water quality through the use of spaceborne Earth Observation (EO) data is a potential

Proceedings of the Eighth International Conference on Environmental Management, Engineering, Planning & Economics
Thessaloniki, Greece, July 20-24, 2021
ISBN: 978-618-5494-53-7

solution to the obstacles of in situ monitoring [5–7]. Estimating quality features directly from satellite imagery can allow the rapid identification of water bodies that suffer from qualitative problems in a more effective and efficient manner [6]. Empirical and analytical approaches can be used to retrieve water constituents. Analytical approaches utilize inherent and apparent optical characteristics, independent of the available ground truth data and retain greater transferability [8]; however, this approach is highly reliant on extensive and accurate optical characteristics, and is highly dependent on the availability and efficacy of the atmospheric correction model [9]. Contrary to the prior, empirical approaches do require the use of in situ measurements to establish a relationship between them and the radiance measured in one or more sensor bands, with no interference from underlying atmospheric. These models can be generated through: statistical regression with band ratios, single/multiple band algebra, Artificial Neural Networks (ANN), Support Vector Machine (SVM), Particle Swarm (PSO), Genetic programming (GP) and unsupervised/supervised classification. Most commonly used models are those generated with linear regression techniques, as they often present reliable results when applied on sites where they were generated. However, their accuracy decreases when applied to other water bodies, due to the changing site-to-site nature of the components of the water bodies [8].Focusing on empirical models, water quality regression estimation models were generated to showcase the capability of application of satellite remote sensing in the characterization of inland lake water quality features at the foothills of Kerkini Mountain in northern Greece. Models were generated using the relationship between temperature, Secchi disk depth, pH, dissolved oxygen, electric conductivity and nutrients (nitrates and phosphorus), and the reflectance recorded in Landsat ETM+ and OLI.

2. MATERIALS AND METHODS

2.1 Study area

Lake Kerkini (41°13’N, 23°08’E) refers to the artificial lake (reservoir) constructed in 1932 and the surrounding wetland area (Figure 1). Its surface area, according to the height of the water-surface, ranges from 37.34 km^2 to 72.52km^2. Lying at the transboundary area of Strymonas River in northern Greece close to the border with Bulgaria, Lake Kerkini’s drainage area extends over 11,600km^2, with the Hellenic sub-basin making up to 803 km^2 [10,11]. Kerkini climate is an intermediate between Mediterranean and Mid-European, with hot summers and cold winters, and average annual rainfall reaching 463.5 mm [12]. Lake Kerkini has developed into one of the most popular stops for migratory bird populations in Europe, as well a wetland of international significance; established as a Natura 2000 protected area and a RAMSAR wetland of international importance. Kerkini accommodates over 300 bird species; with at least 1300 plant species; including indigenous and rare species, as well as Greece’s largest water buffalo population (*Bubalus bubalis*). Deforestation and cultivation along the river corridor has caused extreme erosion of the slopes, the eroded materials having been transported to the Lake. Additionally high concentrations of artificial polluting materials are brought into the lake system by Strymonas River. Thus, monitoring lake water quality is crucial in order to support the Lake Kerkini Management Authority (LKMA) with an enhanced monitoring capacity to protect the biodiversity and the species that depend from the lake, following the guidelines of the Water Framework Directive 2000/60/EC

2.2 Water quality samples

Lake Kerkini Management Authority carried out water quality measurements on a monthly basis between 2010 and 2015 along seven sites (Vironeia, Center of Kerkini Lake, Kerkinitis, Lithotopos, Mesaia, Bistritsa, and Noufara), two of which (Lithotopos and Mesaia) had no samples after January 2012. Sampled water quality parameters included temperature (T, °C), pH, electric conductivity (EC, μS/cm) and dissolved oxygen (DO, mg/L) collected using a CONSORT C392 instrument. Secchi disk depth (SDD, cm), was collected with a handheld Secchi disk Instrument. Water samples were analysed using HANNA HI 83200 Multiparameter Photometer [13], to extract the concentrations of phosphate (PO_4, mg/L) and nitrate (NO_3, mg/L).

2.3 Remote Sensing imagery

Landsat's predominant use in environmental management as well as 40 years of calibrated high spatial resolution data makes it suitable in the study to monitor water quality parameters. Landsat 7 (ETM+) and Landsat 8 (OLI/TIRS) satellite images, were acquired from the USGS Global Visualization Viewer (GloVis), and utilized for suggesting the best suited workflows to monitor lake water quality status. The selected images (path 184; row 31, path 183 row 31 and path 183; row 32) were acquired with up to seven day difference from in situ collection date. This margin was adopted in order to increase the availability of images with cloud cover less than 15%. In addition for certain parameters (*i.e.* SDD), longer windows up to seven days may be acceptable [14,15].

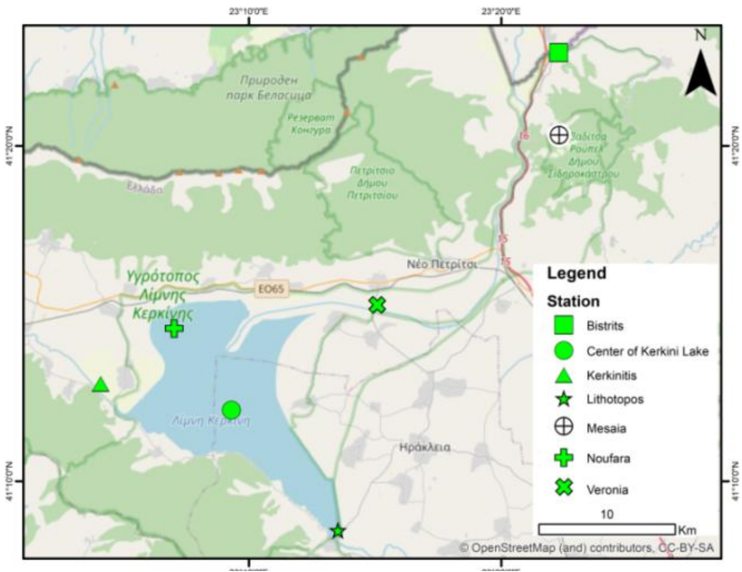

Figure 1: *Location of in situ sampling stations in Kerkini Lake*

2.4 Image processing methodology

The image processing methodology included radiometric and atmospheric corrections, extraction and averaging of spectral reflectance band values at each sample point and statistical analysis (regression). The first two stages were completed using ENVI "Environment for Visualizing Images" software version 5.5. The metafiles of each Landsat image, containing the gains, offset, solar irradiance, sun elevation and acquisition time, were

ISBN: 978-618-5494-53-7

associated in the conversion of digital numbers (DN) to Top-of-Atmosphere (TOA) reflectance. This conversion step is essential in the comparison of multiple images for temporal analysis [16,17]. To ensure minimal interference of atmospheric components (i.e. water vapor and aerosols) and accurately determine surface reflectance data, atmospheric correction through dark object subtraction (DOS) is employed by subtracting a pixel value that represents a background signature from each of the bands [18,19].

Three out of the seven locations (Center of Kerkini Lake, Lithotopos and Noufara) are considered, due to their location in the central portion of the inland water, where reflectance would not be affected by vegetation, shoreline or the reservoir floor [20]. To extract reflectance data for analysis, a 3x3 pixel window centered on the x, y location of the in situ sample point, is selected. The selection of a nine-pixel window is driven by the assumption that water is heterogeneous and often in flux, due to seasonal, solar, and meteorological factors [14,20], and to account for the possibility of spatial misregistration between the GPS position of the water sampling position and the respective Landsat pixel position [14]. The extracted 3x3 pixel windows are resampled to the average over the 90 m pixel area for each band. These values are then combined with the consequent Landsat image date based on a time window (*i.e.*, the time before and after sampling event) of 7 days.

2.5 Models for estimating water quality parameters

Preliminary testing between step-wise multiple linear regression, partial least square regression, simple linear regression and random forest regression had shown that step-wise multiple linear regression achieved the highest accuracies. Step-wise multiple linear regression (SWLR) was performed using the in situ parameter measures as dependent variable, the collection of reflectance bands as the independent variables and the selected level of significance of 0.05. These comprise of: 1) single bands, 2) single band ratio and 3) product of paired bands (Table 1). Model estimation was done on three different time frames (date ranges between image acquisition and in situ sampling): i) +/- 1 day, ii) +/- 3 days and iii) +/- 7 days. Parameter model estimation for ETM+ and OLI/TIRS sensors datasets were handled separately, for each time frame, taking the calibration dataset as the data in the specific time frame and validated against the rest of the available data. In the case of the +/- 7 data frame, validation was based on leave-one-out cross-validation. In order to generate parameter models that could be utilized for either Landsat ETM+ or OLI/TIRS satellites (known hereafter as Landsat common), step-wise multiple linear regression was employed in the same manner, taking into account a collection of bands from both sensors. The selected common bands are represented as those for ETM+ and OLI/TIRS, respectively: Blue band (B_1, B_2), Green band (B_2, B_3), Red band (B_3, B_4), NIR (B_4, B_5), SWIR-1 (B_5, B_6), SWIR-2 (B_7, B_7) and TIR (average of B_{6H} and B_{6L}, average of B_{10} and B_{11}). The in situ and imagery data from overlapping periods (same in situ dates covered by ETM+ and OLI/TIRS) were used as calibration data, and the non-overlapping sample datasets were used as validation dataset.

Several candidate step wise multiple linear regression equations (empirical models) were generated for each water quality feature. The coefficient of determination (R^2) and root mean square error (RMSE) were calculated. To assess capacity in predicting parameters values, these equations were applied to all available data, and the accuracy between the in situ measurements and predicted measurements was calculated (predicted accuracy - PA) for each feature. Additionally, to assess equation transferability from ETM+ to OLI/TIRS and vice versa, bands in Landsat ETM+ equations were replaced to their similar bands in OLI/TIRS equations, and the accuracy (accuracy of transferring - AT) between the predicted values and in situ measurements were calculated as well. As for finding empirical models with ability to be used across Landsat sensor, they were applied to the non-overlapping image dates (*i.e.* image window were both ETM+ and OLI/TIRS are different), where the accuracy (merged data accuracy - MDA) is calculated between the measured samples and the predicted samples in the non-overlapping dates.

Table 1: Collection of bands used as independent variables in the regression analysis for water quality features' derivation from ETM+, OLI/TIRS and Landsat common bands

	Single Bands	*Band Ratios*	*Paired band product*
ETM+	B_1, B_2, B_3, B_4, B_5, B_7, B_{6L}, B_{6H}	B_1/B_2, B_1/B_3, B_1/B_4, B_1/B_5, B_1/B_7, B_2/B_1, B_2/B_3, B_2/B_4, B_2/B_5, B_2/B_7, B_3/B_1, B_3/B_2, B_3/B_4, B_3/B_5, B_3/B_7, B_4/B_1, B_4/B_2, B_4/B_3, B_4/B_5, B_4/B_7, B_5/B_1, B_5/B_2, B_5/B_3, B_5/B_4, B_5/B_7, B_7/B_1, B_7/B_2, B_7/B_3, B_7/B_4, B_7/B_5, B_{6L}/B_{6H}, B_{6H}/B_{6L}	B_1^2, B_2^2, B_3^2, B_4^2, B_5^2, B_7^2, B_{6L}^2, B_{6H}^2, B_1*B_2, B_1*B_3, B_1*B_4, B_1*B_5, B_1*B_7, B_2*B_3, B_2*B_4, B_2*B_5, B_2*B_7, B_3*B_4, B_3*B_5, B_3*B_7, B_4*B_5, B_4*B_7, B_5*B_7, $B_{6H}*B_{6L}$
OLI/ TIRS	B_2, B_3, B_4, B_5, B_6, B_7, B_{11}, B_{12}	B_2/B_3, B_2/B_4, B_2/B_5, B_2/B_6, B_2/B_7, B_3/B_2, B_3/B_4, B_3/B_5, B_3/B_7, B_4/B_1, B_4/B_2, B_4/B_3, B_4/B_5, B_4/B_6, B_4/B_7, B_5/B_2, B_5/B_3, B_5/B_4, B_5/B_6, B_5/B_7, B_6/B_2, B_6/B_3, B_6/B_4, B_6/B_5 B_6/B_7, B_7/B_2, B_7/B_3, B_7/B_4, B_7/B_5, B_7/B_6, B_{10}/B_{11}, B_{11}/B_{10}	B_2^2, B_3^2, B_4^2, B_5^2, B_6^2, B_7^2,B_{10}^2, B_{11}^2 B_2*B_3, B_2*B_4, B_2*B_5, B_2*B_6, B_2*B_7, B_3*B_4, B_3*B_5, B_3*B_6, B_3*B_7, B_4*B_5, B_4*B_6, B_4*B_7, B_5*B_6, B_5*B_7, B_6*B_7, $B_{10}*B_{11}$
ETM+ or OLI/ TIRS common bands	Blue, Green, Red, NIR, SWIR1, SWIR2, TIR	Blue/Green, Blue/Red, Blue/NIR, Blue/SWIR1, Blue/SWIR2, Green/Blue, Green/Red, Green/NIR, Green/SWIR1, Green/SWIR2, Red/Blue, Red/Green, Red/NIR Red/SWIR1, Red/SWIR2, NIR/Blue, NIR/Green, NIR/Red, NIR/SWIR1, NIR/SWIR2, SWIR1/Blue, SWIR1/Green, SWIR1/Red, SWIR1/NIR, SWIR1/SWIR2, SWIR2/Blue, SWIR2/Green, SWIR2/Red, SWIR2/NIR, SWIR2/SWIR1	$Blue^2$, $Green^2$, Red^2, NIR^2, $SWIR1^2$, $SWIR2^2$, TIR^2, Blue*Green, Blue*Red, Blue*NIR, Blue*SWIR1, Blue*SWIR2, Green*Red, Green*NIR, Green*SWIR1, Green*SWIR2, Red*NIR, Red*SWIR1, Red*SWIR2, SWIR1*SWIR2

3. RESULTS AND DISCUSSION

Following the processing of the images and utilizing a 3x3 pixel window for each in situ sampling station location [14,20] the regression model with the highest R^2, lowest RMSE, and highest predictive capability from cross-validation was identified. The possibility to apply derived equations to other sensors was also examined (see Tables 2 and 3).

3.1 Estimation of water surface temperature

The water surface temperature empirical model estimation for Landsat ETM+, OLI/TIRS and Landsat common, included the presence of the thermal band. Water surface temperature estimation for ETM+ (Table 2) during a time frame of +/- 1 day away from image the acquisition date explained 0.99 (R^2) of the variance, with the accuracy resulting from predicting values against measured values in the larger time frame dataset being PA = 82.5%. This equation included the bands B_1, B_2, B_3 and B_{6L}. When applied to the OLI/TIRS dataset (Table 2) it achieved an AT = 63.5% between the in situ values and the predicted ones.

Estimating water surface temperature with OLI/TIRS band collection included bands: B_{11}, B_{11}^2 and the product of B_4*B_7. This equation explained 0.94 (R^2) of the variance in a time frame of +/-3 days away from image acquisition, while achieving a PA = 93.8% when validating the model prediction accuracy against the larger dataset. Once this OLI/TIRS model was applied to Landsat ETM+ dataset, it was able to predict the measured values with an AT = 86%. The equation derived to be used across Landsat sensors (Table 3), for estimating water surface temperature contained only the thermal TIR band and explained 0.84 (R^2) of the variance in the +/-3 days dataset. This empirical estimation model achieved a PA = 79.6% when applied to the larger dataset (+/- 7 day timeframe), and also achieved MDA = 82.2%, when estimating temperature in the non-overlapping dataset.

The incorporation of thermal bands, B_{6L}/B_{6H} (Landsat ETM+), B_{10}/B_{11} (Landsat OLI/TIRS) and TIR (Thermal band of the Landsat common bands) has been previously successful in the estimation of surface temperature [16,21], as it has also been proven in this study. The SWIR band (B_7), included into the OLI/TIRS temperature estimation model, is inversely proportional to moisture [22,23], since an increase in water content decreases temperature. Likewise, green (B_2, ETM+ - Table 2) and red (B_3, ETM+; B_4, OLI/TIRS - Table 2) bands, used to measure green reflectance peaks of vegetation and chlorophyll absorption regions, respectively, are linked with the presence of vegetation, which also decreases temperature [23]. These models can estimate water surface temperature even if validation of samples are 7 days away from the image acquisition date.

3.2 Estimation of Secchi disk depth

Secchi disk depth is an attribute of water transparency. Performing step-wise multiple linear regression using the Landsat ETM+ data, resulted in no significant empirical models capable of explaining more than 0.50 of the variation (Table 2). The Landsat OLI/TIRS model, employed a low number of samples (n = 6) in the time frame of +/- 1 day away from image acquisition and was able to explain (R^2) 0.99 of the variation. However, this low sample size caused an overfitting of the equation to the dataset, thereby affecting the accuracy of assessment of the SDD equation [24], as it achieved a PA = 26.3%, when applied to the larger dataset (Table 2). The equation derived to estimate SDD across the Landsat sensors (Table 3) incorporated ratios of Blue/NIR, Green/Red and SWIR-2/Red and the product of NIR*SWIR2. This combination of bands was able to explain the variation with an R^2 = 0.88. Its capacity when predicting each value in the whole dataset through the leave-one-out cross-validation process resulted in a PA = 78%. When this equation estimated SDD against the measured in situ values in the non-overlapping region, it resulted in a MDA = 42%.

The results show that more samples for each image analyzed can provide accurate and usable equations for SDD estimation [25,26]. In addition to their small number, the in situ measurements themselves, can be affected by the subjective observation of the person taking the measurement, the solar angle and the margin of error of the instrument [27]. The blue (Blue, Landsat common- Table 3), green (B_3, OLI/TIRS-Table 2, Green, Landsat common-Table 3), red (B_4, OLI/TIRS-Table 2, Red, Landsat common-Table 3) [6,28], NIR and SWIR-2 (Landsat common-Table 3) [24] bands, have been incorporated in the SDD estimation equations in the past. However, the over-fitting of the equations to the dataset, due to the small sample size, hinders our ability to estimate SDD from Landsat imagery.

Table 6: Empirical models derived from Landsat ETM+ and Landsat OLI/TIRS for estimating water quality features. These equations explain the variation with an $R^2 > 0.50$. These parameters include: temperature (T, °C), pH (unitless), electric conductivity (EC, μS/cm), dissolved oxygen (DO, mg/L), Secchi disk depth (SDD, cm), phosphate (PO_4, mg/L) and nitrate (NO_3, mg/L). PA is the accuracy between in situ measurements and predicted measurements. AT is the accuracy of predicting feature values when transferring amongst sensors.

Parameter and equation: Landsat ETM+	*Time frame (+/- days)*	R^2	*RMSE*	*PA*	*AT*
$T = -9.92 + 104.4\times B_1 + 193.6\times B_2 - 286.3\times B_3 + 1.1899\times B_{6L}$	1	0.99	0.90	82.5	63.5
$pH = 8.206 + 0.04873 \times B_{6H}$	3	0.52	0.41	37.6	60.8
$DO = 16.37 - 16.21\times(B_3\times B_5)$	1	0.60	3.73	21.2	15.0
$EC = 520.8 - 11.18\times B_{6H}$	1	0.56	17.79	43.7	73.9
$PO_4 = 3.735 - 2.762\times B_1^2 - 0.022\times(B_5/B_7)$	1	0.95	2.23	0.00	58.2
$NO_3 = -0.511 + 0.295\times (B_3/B_4) + 0.66\times(B_4/B_3) - 0.035\times B_{6H} + 0.001\times B_{6L}^2$	7	0.79	0.2	69.8	4.7
Parameter and equation: Landsat OLI/TIRS	*Time frame (+/- days)*	R^2	*RMSE*	*PA*	*AT*
$T = 0.787 + 1.124\times B_{11} - 0.009\times B_{11}^2 + 6.071\times(B_4\times B_7)$	3	0.94	2.25	93.8	86.0
$SDD = 155.495 - 203.145\times(B_3\times B_4) + 2039.598\times B_6$	1	0.99	2.35	26.3	21.3
$pH = 7.491 + 0.0476\times B_{10}$	7	0.62	0.36	58.1	37.8
$DO = 15.189 - 4.190\times(B_3/B_2) - 0.002\times B_{11}^2 - 2.133\times(B_4\times B_7)$	3	0.85	0.75	10.2	1.6
$EC = 340.9 + 480\times B_2 - 5.479\times B_{10}$	3	0.81	35.12	84.2	44.2
$NO_3 = 3.17 - 0.1144\times B_{11}$	3	0.91	0.29	47.1	1.0

3.3 Estimation of water acidity

Performing step-wise multiple linear regression using Landsat ETM+, OLI/TIRS and Landsat common, resulted in equations capable of estimating water acidity, represented by pH. The use of only a thermal band, in the linear formula of: a*(Thermal Band) + b, in ETM+ and OLI/TIRS empirical models (Table 2), explained 0.52 and 0.62 of the variation in a time frame of +/- 3 days and +/- 7 day away from image acquisition respectively (Table 2). The ETM+ equation, applied to the larger dataset (+/- 7 days), was able to predict the pH values with a PA = 37.6%, whereas the OLI/TIRS equation, achieved a PA = 58.1%. The ETM+ equation when applied to OLI/TIRS dataset estimated pH with an AT = 60.8%, while the OLI/TIRS equation applied to the ETM+ dataset predicted pH values with an AT = 37.8%. The equation for estimating pH across Landsat sensors (Table 3), included, the Red and NIR bands, in addition to the thermal band, and explained the variation of pH in Kerkini Lake with an $R^2 = 0.82$ in the time frame of +/-7 days, as well as achieving a lower RMSE than the other equations. Its capacity in predicting each value in the +/- 7 day timeframe, had a PA = 72.5%, while using it to assess the accuracy between the measured and predicted in situ values in the non-overlapping region, resulted in an accuracy MDA = 36.2%.

Using empirical models in pH estimation has not been thoroughly addressed in literature [29,30]. Estimating pH depends mostly on the availability of in situ samples. Thermal bands (B_{6H}, ETM+; B_{10}, OLI/TIRS; TIR, Landsat common - see Tables 2 and 3) included in the empirical models, are linked to the hydrogen ionization process, where an increase in temperature favors hydrogen ionization in water thus lowering pH values. NIR and Red bands, incorporated into equation to be used across Landsat sensors (Table 3), can be linked to pH indirectly, by observing phytoplankton blooms, which limit pH values or an increase in dissolved carbon concentrations which tends to increase pH values [6,31].

ISBN: 978-618-5494-53-7

Table 3: Empirical models derived with the ability to be used across Landsat sensor for estimating water quality features. These equations explain the variation with an $R^2 > 0.50$. These parameters include: temperature (T, °C), pH (unitless), electric conductivity (EC, μS/cm), dissolved oxygen (DO, mg/L), Secchi disk depth (SDD, cm), phosphate (PO_4, mg/L) and nitrate (NO_3, mg/L). PA is the accuracy between in situ measurements and predicted measurements. MDA is the accuracy of predicting feature values when applied to the non-overlapping image dates of Landsat ETM+ and OLI/TIRS.

Parameter and equation: Landsat common	*Time frame (+/- days)*	R^2	*RMSE*	*PA*	*MDA*
$T = 5.25 + 0.871 \times TIR$	3	0.84	5.51	79.6	82.2
$SDD = -115.8 + 13.05 \times (Blue/NIR) + 53.3 \times (Green/Red) + 98.7 \times (SWIR2/Red) + 53.95 \times (NIR \times SWIR2)$	7	0.88	0.84	78.0	42.0
$pH = 8.879 - 25.3 \times Red + 23 \times NIR + 0.023 \times TIR$	7	0.82	0.32	72.5	36.2
$DO = 7.98894 + 3.0449 \times Red^2$	1	0.99	0.35	7.0	2.3
$EC = 357.4 + 560 \times Blue - 5.336 \times TIR - 84.9 \times (SWIR1/Blue)$	7	0.80	29.0	70.8	51.1
$PO_4 = 3.340 - 1.454 \times (SWIR2/Green) - 0.0017 \times TIR - 2.914 \times (Green \times NIR)$	3	0.95	1.07	36.0	15.3
$NO_3 = 0.909 + 0.366 \times (SWIR2/Blue) - 1.695 \times SWIR1^2 - 0.63 \times SWIR2^2 - 0.88 \times (Blue \times Red) + 2.118 \times (Red \times NIR)$	7	0.89	0.38	61.8	6.2

3.4 Estimation of dissolved oxygen

The empirical models estimating dissolved oxygen (DO) from Landsat ETM+, OLI/TIRS and Landsat common, explained its variation with an $R^2 > 0.59$ in Kerkini Lake. The equation resulting from SWLR using ETM+ satellite data (Table 2) incorporated the product of the Red band (B_3) and SWIR-1 (B_5) with an image window of +/-1 day. Its accuracy in predicting DO values in the larger dataset against in situ measurements was PA = 21.2%, while its application to OLI/TIRS dataset predicted the values with an AT = 15%. The equation resulting from using OLI/TIRS satellite (Table 2) explained the DO variation with an $R^2 = 0.85$. It incorporated the Green (B_3) to Blue (B_2) ratio, the product of Red (B_4) and SWIR-2 (B_7) and the squared product of the thermal band (B_{11}), in an image window of +/-3 days. Very low accuracies resulted from applying the equation to predict DO values in a larger dataset (+/- 7 day timeframe) (PA = 10.2%), as well as predicting DO values in the Landsat ETM+ dataset (AT = 1.7%). The equation for estimating DO across Landsat sensors (Table 3), included the squared product of the Red band and explained the variation of DO in Kerkini Lake with an $R^2 = 0.999$ in a time frame of +/-1 days. Its accuracy between predicting dissolved oxygen vs measured in situ values, achieved a PA = 7% when applied to the larger dataset, and a MDA = 2.3%, when applied to the non-overlapping dates between ETM+ and OLI/TIRS. These low predictive capabilities limit the ability to apply these models for estimating DO.

This poor accuracy for DO concentration retrieval is not uncommon in the literature [29,30,32], especially with a small sample size. Thermal and red bands have been shown to have the ability to estimate dissolved oxygen. The incorporation of thermal bands helped achieve high predictive accuracies in estimating dissolved oxygen concentration in water [32], as temperature and dissolved oxygen are inversely proportional. The red band, an attribute of vegetation linked to the photosynthetic process, (Table 3), didn't provide any promising results when included in the equation derived for estimating DO concentration across Landsat sensors. The small sample size of in-situ data, resulted in the overfitting of the model to the dataset, which was evident when predicting the values in the +/- 7 image window (PAs for all three empirical models were < 20%).

3.5 Estimation of electric conductivity

Electric conductivity is a measure related to the concentration of salts present in the water. The ETM+, OLI/TIRS and the Landsat common empirical estimation models, estimated the electric conductivity in Kerkini lake with an R^2 = 0.55, 0.81 and 0.80, respectively (Table 2, 3). The best-fit derived ETM+ equation, incorporated the thermal band (B_{6H}) in an image window of +/-1 day. Its application to the larger dataset of ETM+, resulted in an accuracy in predicting EC values against the in situ measurement with a PA = 43.7%, while its estimation of EC in the OLI/TIRS dataset resulted in an accuracy of TA = 73.9%. The combination of the blue band (B_2) with the thermal band (B_{10}) in the OLI/TIRS equation increased the explanation of variability of DO to R^2= 0.81, with an accuracy of PA = 84.2% in predicting the values in the larger dataset . However, once this empirical model was applied to estimate ETM+ EC values, it achieved an accuracy of TA = 44.2% between the predicted and measured EC values. The empirical model for estimating EC across Landsat sensors (Table 3), included the SWIR1 band, in addition to the previous bands in the other equations. This resulted in the equation explaining R^2 = 0.80 of the variability of EC. Its capacity in predicting each value in the +/- 7 day timeframe, through leave-one-out cross validation achieved a PA = 70.8%, while using it to assess the accuracy between the measured and predicted in situ values in the non-overlapping region, resulted in an accuracy MDA = 51.1%.

The incorporation of mainly the blue (B_2, OLI/TIRS – Table 2, Blue, Landsat common – Table 3) and/or thermal bands for estimating electric conductivity in water, increases the accuracy of estimation, and the applicability of use of the model [22]. The relationship between electric conductivity and temperature in electrolyte solutions is primarily controlled by the viscosity-temperature relation of pure water, where a 2% increase of electric conductivity coincides with a one-degree Celsius increase of temperature [33]. These empirical estimation models are applicable, where validation of the models through in situ sampling can be done within 3 days before or after satellite acquisition.

3.6 Estimation of nutrients (Phosphate and Nitrate)

The nutrients, phosphorus and nitrate, are important biological growth and eutrophication indicators, playing an important part in water quality assessment. There has been no consensus on the appropriate bands that define the estimation of both parameters. The best-fit equation using ETM+ satellite data (Table 2), explained the variation of phosphate with an R^2 = 0.95. This empirical model was applied to the larger dataset (+/-7 day timeframe) and to the OLI/TIRS dataset. It predicted the phosphate values with an accuracy PA = 0% and TA = 58.2%, against the in situ measurements. This 0% accuracy is due to the small sample size (Table 2). Performing SWLR using the OLI/TIRS data, resulted in no empirical model with the ability to explain the variation in Kerkini Lake with an $R^2 > 0.5$. The equation for estimating phosphate across Landsat sensors, explained the variation in phosphate concentration with an R^2=0.95 in a +/-3 image window (Table 3). It's accuracy in predicting phosphate against in situ measurements in a larger dataset and in the dates where ETM+ and OLI/TIRS don't overlap, resulted in a PA = 36.0% and MDA = 15.3%, respectively. Certain studies have shown that incorporating the blue band (B_1, Landsat ETM+, see Table 2) [34,35] and the SWIR bands (Landsat common, see Table 3) may help in estimating phosphorus [36]. Additionally, indirect indications of phosphorus concentrations could be related to the thermal band (i.e. TIR, see Table 3), which may reveal surface roughness information, however, this has not been concluded yet [37,38].

ISBN: 978-618-5494-53-7

Estimating nitrate concentration using the ETM+ dataset (Table 2), resulted in an equation explaining the variation of nitrate with an R^2 = 0.79, in a +/-7 day timeframe, achieving a PA = 69.8% when validated, through a leave-one-out cross validation. However, this equation applied on the OLI/TIRS data, achieved an accuracy between the predicted values and in situ measurements of TA = 4.7%. Estimating nitrate concentrations through the use of the OLI/TIRS dataset resulted in an empirical model, containing only the thermal band (B_{11}), while explaining R^2 = 0.91 of the variation of nitrate in Kerkini Lake in a timeframe of +/- 3 days (Table 2). The ETM+ equation, applied to the larger dataset (+/- 7 days), was able to predict the nitrate concentration with a PA = 47.1% against in situ measurements, while applying it to the ETM+ dataset, achieved an accuracy AT = 1.0% between the measured and predicted values. Estimating nitrate across the Landsat sensors, was through an SWLR equation that explained the variation of nitrate concentration with a R^2=0.89 in a +/-7 day timeframe. This equation included the combination of the blue, red, NIR and SWIR-1 and SWIR-2 bands. Its capacity in predicting each value in the +/- 7 day timeframe, through the leave-one-out cross validation achieved a PA = 61.8%, while using it to assess the accuracy between the measured and predicted in situ values in dates where ETM+ and OLI/TIRS don't overlap, resulted in an accuracy MDA = 6.2%. Studying these nutrients in literature has been avoided and primarily focused on trend analysis rather than estimation and modeling [37].

4. CONCLUSIONS

Remote sensing can support the monitoring and policy implementation of water quality measures in lakes, especially for the development of protection measures and management practices. The results of this research show that Landsat imagery (ETM+ and OLI/TIRS) data can be used to perform water quality features' estimation in Kerkini Lake. Temperature, pH and electric conductivity empirical models can be used across sensors (Landsat ETM+ or OLI/TIRS), even if the calibration for these models (in situ measurements) is carried out at +/- 3 days away from image acquisition. For these variables, moderate to high accuracies are registered. The low volume of samples available for this study, hindered the estimation of SDD. On the other hand, the estimation of DO and nutrients has not been sufficiently accurate, which is something that has also been observed in the literature. In general, a higher number of in situ measurements and sampling sites would be required to reach to statistically reliable results.

Acknowledgements

The authors wish to express their gratitude to the personnel of the Lake Kerkini Management Authority for providing access to the databases with water quality data, specifically Th. Nazaridis and M. Davis for the analysis of water quality samples. This work has been partially funded and supported by the European Union's Horizon 2020 innovation program under Grant Agreement No. 820852, Eshape (https://e-shape.eu/).

References

[1] Millennium Ecosystem Assessment, 2005. Ecosystems and Human Well-being: Synthesis.
[2] Kachroud, M., Trolard, F., Kefi, M., Jebari, S., and Bourrié, G., 2019. Water quality indices: Challenges and application limits in the literature. *Water*, **11** (**2**), 1–26.
[3] Darwish, T., Atallah, T., and Fadel, A., 2018. Challenges of soil carbon sequestration in NENA Region. *SOIL*, **4**, 225–235.

[4] Fadel, A., Kanj, M., and Slim, K., 2021. Water Quality Index variations in a Mediterranean reservoir: a multivariate statistical analysis relating it to different variables over 8 years. *Environmental Earth Sciences*, **80** (**2**), 65.

[5] Dörnhöfer, K., and Oppelt, N., 2016. Remote sensing for lake research and monitoring – Recent advances. *Ecological Indicators*, **64**, 105–122.

[6] Gholizadeh, M., Melesse, A., and Reddi, L., 2016. A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors*, **16** (**8**), 1298.

[7] Palmer, S.C.J., Kutser, T., and Hunter, P.D., 2015. Remote sensing of inland waters: Challenges, progress and future directions. *Remote Sensing of Environment*, **157**, 1–8.

[8] Chang, N.-B., Imen, S., and 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.

[9] Pavelsky, T.M., and Smith, L.C., 2009. Remote sensing of suspended sediment concentration, flow velocity, and lake recharge in the Peace-Athabasca Delta, Canada. *Water Resources Research*, **45** (**11**).

[10] Ovakoglou, G., Alexandridis, T.K., Crisman, T.L., Skoulikaris, C., and Vergos, G.S., 2016. Use of MODIS satellite images for detailed lake morphometry: Application to basins with large water level fluctuations. *International Journal of Applied Earth Observation and Geoinformation*, **51**, 37–46.

[11] Psilovikos, A., and Margoni, S., 2010. An empirical model of sediment deposition processes in Lake Kerkini, Central Macedonia Greece. *Environmental Monitoring and Assessment*, **164** (**1–4**), 573–592.

[12] Gerakes, P.A. (ed.), 1989. Conservation and Management of Greek Wetlands. *Proceedings of a Workshop on Greek Wetlands*, 493.

[13] Kerkini Lake Management Authority, 2018. ΣΧΕΔΙΟ ΔΙΑΧΕΙΡΙΣΗΣ ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ. Translation: Water Resources Management Plan.

[14] Barrett, D., and Frazier, A., 2016. Automated Method for Monitoring Water Quality Using Landsat Imagery. *Water*, **8** (**6**), 257.

[15] Kloiber, S.M., Brezonik, P.L., and Bauer, M.E., 2002. Application of Landsat imagery to regional-scale assessments of lake clarity. *Water Research*, **36** (**17**), 4330–4340.

[16] Tavares, M., Cunha, A., Motta-Marques, D., Ruhoff, A., Cavalcanti, J., Fragoso, C., Martín Bravo, J., Munar, A., Fan, F., and Rodrigues, L., 2019. Comparison of Methods to Estimate Lake-Surface-Water Temperature Using Landsat 7 ETM+ and MODIS Imagery: Case Study of a Large Shallow Subtropical Lake in Southern Brazil. *Water*, **11** (**1**), 168.

[17] Young, N.E., Anderson, R.S., Chignell, S.M., Vorster, A.G., Lawrence, R., and Evangelista, P.H., 2017. A survival guide to Landsat preprocessing. *Ecology*, **98** (**4**), 920–932.

[18] Chavez, P.S., 1996. Image-based atmospheric corrections - Revisited and improved. *Photogrammetric Engineering and Remote Sensing*, **62** (**9**), 1025–1036.

[19] Sharaf El Din, E., 2019. Enhancing the accuracy of retrieving quantities of turbidity and total suspended solids using Landsat-8-based-principal component analysis technique. *Journal of Spatial Science*, 1–20.

[20] Hadjimitsis, D.G., and Clayton, C., 2009. Assessment of temporal variations of water quality in inland water bodies using atmospheric corrected satellite remotely sensed image data. *Environmental Monitoring and Assessment*, **159** (**1–4**), 281–292.

[21] Lamaro, A.A., Mariñelarena, A., Torrusio, S.E., and Sala, S.E., 2013. Water surface temperature estimation from Landsat 7 ETM+ thermal infrared data using the generalized single-channel method: Case study of Embalse del Río Tercero (Córdoba, Argentina). *Advances in Space Research*, **51** (**3**), 492–500.

[22] Khattab, M.F.O., and 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.

[23] Lillesand, T.M., Kiefer, R.W., and Chipman, J.W., 2015. *Remote Sensing and Image Interpretation*, John Wiley & Sons, Inc.
[24] Brezonik, P., Menken, K.D., and Bauer, M., 2005. Landsat-based remote sensing of lake water quality characteristics, including chlorophyll and colored dissolved organic matter (CDOM). *Lake and Reservoir Management*, **21** (**4**), 373–382.
[25] Olmanson, L.G., Brezonik, P.L., Finlay, J.C., and Bauer, M.E., 2016. Comparison of Landsat 8 and Landsat 7 for regional measurements of CDOM and water clarity in lakes. *Remote Sensing of Environment*, **185**, 119–128.
[26] Sriwongsitanon, N., Surakit, K., and Thianpopirug, S., 2011. Influence of atmospheric correction and number of sampling points on the accuracy of water clarity assessment using remote sensing application. *Journal of Hydrology*, **401** (**3**–**4**), 203–220.
[27] Cox, R.M., Forsythe, R.D., Vaughan, G.E., and Olmsted, L.L., 1998. Assessing Water Quality in Catawba River Reservoirs Using Landsat Thematic Mapper Satellite Data. *Lake and Reservoir Management*, **14** (**4**), 405–416.
[28] Wu, C., Wu, J., Qi, J., Zhang, L., Huang, H., Lou, L., and Chen, Y., 2010. Empirical estimation of total phosphorus concentration in the mainstream of the Qiantang River in China using Landsat TM data. *International Journal of Remote Sensing*, **31** (**9**), 2309–2324.
[29] Abdullah, H.S., 2016. Water Quality Assessment Models for Dokan Lake Using Landsat 8 OLI Satellite Images.
[30] Mushtaq, F., and Nee Lala, M.G., 2017. Remote estimation of water quality parameters of Himalayan lake (Kashmir) using Landsat 8 OLI imagery. *Geocarto International*, **32** (**3**), 274–285.
[31] Park, E., and Latrubesse, E.M., 2014. Modeling suspended sediment distribution patterns of the Amazon River using MODIS data. *Remote Sensing of Environment*, **147**, 232–242.
[32] Arief, M., 2017. Development of Dissolved Oxygen Concentration Extraction Model using Landsat Data Case Study: Ringgung Coastal Waters. *International Journal of Remote Sensing and Earth Sciences (IJReSES)*, **12** (**1**), 1.
[33] Hayashi, M., 2004. Temperature-Electrical Conductivity Relation of Water for Environmental Monitoring and Geophysical Data Inversion. *Environmental Monitoring and Assessment*, **96** (**1**–**3**), 119–128.
[34] Chen, J., and Quan, W., 2012. Using Landsat/TM Imagery to Estimate Nitrogen and Phosphorus Concentration in Taihu Lake, China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **5** (**1**), 273–280.
[35] Kutser, T., Arst, H., Miller, T., Käärmann, L., and Milius, A., 1995. Telespectrometrical estimation of water transparency, chlorophyll-a and total phosphorus concentration of Lake Peipsi. *International Journal of Remote Sensing*, **16** (**16**), 1–2.
[36] Isenstein, E.M., and Park, M.H., 2014. Assessment of nutrient distributions in Lake Champlain using satellite remote sensing. *Journal of Environmental Sciences (China)*, **26** (**9**), 1831–1836.
[37] He, W., Chen, S., Liu, X., and 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.
[38] Zhang, Y., Pulliainen, J., Koponen, S., and Hallikainen, M., 2002. Application of an empirical neural network to surface water quality estimation in the Gulf of Finland using combined optical data and microwave data. *Remote Sensing of Environment*, **81** (**2**–**3**), 327–336.