Using Landsat 8 Imagery for Remote Monitoring of Total Phosphorus as a Water Quality Parameter of Irrigation Ponds in Taiwan

Po-Wen Su
Graduate Institute of Environmental Engineering

Shang-Lien Lo (✉ sllo@ntu.edu.tw)
Graduate Institute of Environmental Engineering  https://orcid.org/0000-0003-1668-6513

Research Article

Keywords: ACOLITE, Forward selection, Multiple regression, Remote sensing, Water quality, Land cover

DOI: https://doi.org/10.21203/rs.3.rs-231078/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Monitoring water body quality parameters with high spatial and temporal resolutions is crucial because mitigation of pollution is usually costlier than early prevention/intervention. The existing monitoring methods for irrigation ponds in Taoyuan, Taiwan, are based on field measurements that have low spatial and temporal resolutions. In this study, using Landsat 8 satellite imagery, a multiple regression-derived relationship between the satellite band reflectance and the concentration of total phosphorus (TP) was established. The satellite imagery was atmospherically corrected with ACOLITE based on shortwave infrared (SWIR) bands. This method was used to select predictor variables in the multiple regression-derived equation based on forward selection of variables using a p-value and variation inflation factor (VIF) threshold. The derived equation yielded a coefficient of determination ($R^2$) of 0.67. The near-infrared band (band 5) was found to be most significant. The Landsat 8 imagery retrieved for two of the three pond studies included only a few pixels from the ponds because parts of the pond surfaces are covered by floating photovoltaic power plants. The TP concentrations resulting from the derived equation indicate the feasibility of using satellite remote sensing methods to monitor the water quality. The derived relationships are potentially applicable to extend the availability of temporal and spatial water quality data for these irrigation ponds.

Introduction

Continuous water quality monitoring for human health and well-being and ecosystems are crucial to survival. Urbanization, agriculture and other human factors can change water quality (Kannel et al., 2007). Once pollution occurs, mitigation measures can be far more costly than early prevention. Although continuous water quality monitoring is important, the cost of adequate temporal and spatial physical measurements can be prohibitive (Harmel et al., 2006).

In recent decades, the increased availability and affordability of satellite imagery has led to its use in monitoring water quality at higher frequencies and lower than traditional methods. Each water quality component exhibits a specific spectral response that satellites can observe. For example, the settlement of suspended sediment usually displays strong light after the spread (Liu et al., 2003), but the actual color depends on the origin of the land (Bukata, 2005). In addition, these findings suggest that multispectral satellite imagery can be used to use various methods to estimate water quality, and most of them use multiple regression analysis or artificial neural networks (ANNs) (Kloiber et al., 2002; Liu et al., 2003; Kishino et al., 2005). In general, visible bands are used to measure water quality because of their ability to penetrate the water column (Liu et al., 2003). However, from the infrared wavelengths (including thermal infrared band), image data are directly used in multiple regression analysis (Barbini et al., 1997) or into the fluid dynamics model (Schott et al., 2001; Pahlevan et al., 2012) to measure the water quality. The applicability of the results is usually limited to the same water body because the spectral response of suspended sediments depends on their terrestrial origin, and the distribution of sediment particle size affects turbidity, even if the concentration of suspended sediments is the same (Liu et al., 2003; Bukata, 2005).
The objective of this study was to use multiple regression analysis on the band reflectance of Landsat 8 imagery to derive the coefficients of an equation for estimating the water quality of irrigation ponds in Taoyuan, Taiwan. Considering the research requirements, Landsat 8 data from the Operational Land Imager (OLI) sensor, which has 9 spectral bands, provide acceptable details (Markham et al., 2014). With no similar studies performed on these irrigation ponds in Taoyuan, this study was needed because water quality equations are often specific to the same body of water and therefore do not apply to equations derived elsewhere. This is the first attempt to derive the relationship between the total phosphorus (TP) concentration, as an index of the water quality, and satellite-derived reflectance for these irrigation ponds. The Landsat data can be used to obtain two water quality estimates monthly, with an optimal number of 24 measurements each year at any location in the ponds; thus, the imagery can be used to extend the availability of the temporal and spatial water quality data for irrigation ponds. Because irrigation ponds are major water bodies in the area, local authorities must promptly detect water abnormalities in them.

**Materials And Methods**

**Study area**

The Taoyuan Tableland, located in northwestern Taiwan, covers an area of 757 km², comprising 2,898 ha of irrigation ponds, and the area has largely been converted to urban land uses due to urbanization and commercialization (Fang et al., 2009). The Taoyuan Tableland is located between the northern boundary of Linkou Tableland (23°05′N, 121°17′E) and the southern boundary of Hukou Tableland (22°55′N, 121°05′E), with the town of Yingge to the east (22°56′ N, 121°20′ E) and the Taiwan Strait to the west (22°75′ N, 120°99′ E) (Department of Land Administration, 2002) (Fig. 1).

In the past, there were no obvious or stable flows of rivers in the Taoyuan area. For residential uses and agricultural irrigation, many water storage facilities, such as irrigation ponds, have been established. In 1913, there were 10,000 ponds, the largest number recorded in the area. With the completion of the Shimen Reservoir and the Taoyuan and Shimen Canals, which provide new sources of irrigation water, many of these ponds no longer provide irrigation functions, but they still retain some other functions, such as decreasing the temperature. There are 2,851 still serving as irrigation ponds in Taoyuan now, according to a recent survey (Chou et al., 2017). Therefore, an effective method for monitoring the irrigation water quality change in ponds to evaluate the effects of urbanization on local water resources will be beneficial.

The Taoyuan Irrigation Association (TIA) and the Shihmen Irrigation Association (SIA) have a number of water quality monitoring stations in this area, as shown in Fig. 1, but only three of them (Shetzu No. 1, Jao No. 32B, and Jao No. 37B) monitored the water quality constituent of interest of this study, namely, the TP, within the time frame of the available satellite images. Additionally, these three stations were relatively new, and recently, floating photovoltaic power plants were placed on some portion of these three ponds, which cover some surfaces of the ponds (Song et al., 2018). The water quality data were based on seasonal field measurements from February 2018 to August 2019, and the data are available.
from the SIA (https://www.smia.gov.tw/water1.asp). Among these three stations, Shetzu No. 1 and Jao No. 37B collected monitoring data on seven days, while Jao No. 32B collected monitoring data only on three days during this period. Jao No. 37B could not be detected by Landsat 8 imagery and was later excluded from this study. Consequently, only Shetzu No. 1 (24°54'25.2"N 121°14'45.0"E) and Jao No. 32B (24°56'17.6"N 121°04'11.1"E) were studied, and their locations are shown in Fig. 1.

Data analysis

Only images with low cloud coverage in the Taoyuan Tableland area were chosen and needed, and this criterion significantly reduced the size of the image pool suitable for analysis. For example, the water quality sampling data collected on February 18, 2019, were excluded from subsequent analysis because the corresponding satellite images within one month were obscured by clouds, as shown in Fig. S1 (Electronic Supplementary Material). The target point was the location of the Taoyuan area in Fig. S1(a)-(d). From the pool of suitable images, five Landsat 8 images collected within one month of water quality measurements obtained by the SIA were selected for this study, as shown in Table 1. All Landsat 8 images were collected from the United States Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center (https://www.usgs.gov/centers/eros). The image specifications of the Operational Land Imager (OLI) spectral bands from Landsat 8 are listed in Table 2. However, this study only used information from band 1 to band 7. Among these bands, band 6 and band 7, so-called shortwave infrared (SWIR) bands, were used for atmospheric correction (Markham et al., 2014), and the others (band 1 to band 5) were processed at a 30-m resolution with ACOLITE (Quinten, 2017). The details from ACOLITE are shown in Fig. S2 and Fig. S3 (Electronic Supplementary Material). A distinct advantage of using the SWIR bands is that it does not require a clear water pixel to determine the type of aerosol. Therefore, imagery is not affected by stray light (Quinten and Ruddick, 2015).

Atmospheric corrections using the SWIR bands can be used to obtain reflectance values without being affected by atmospheric path radiation. (Quinten and Ruddick, 2015). Images from digital number (DN) conversion for spectral radiant (L) first format and then by atmospheric corrections using ACOLITE SWIR bands for processing to determine the surface reflection as

\[ L = M_L \cdot DN + A_L \]  \hspace{1cm} (1)

where \( M_L \) is a multiplicative factor and \( A_L \) is an additive factor provided in the metadata file. L is normalized to the band average of solar irradiance, and the spectral radiation reflectance (\( \rho \)) is calculated as follows:

\[ \rho = \frac{\pi L \cdot d^2}{F_0 \cdot \cos \theta_o} \]  \hspace{1cm} (2)
where $d$ is the distance between the earth and the sun, expressed in astronomical units; $\theta_0$ is the zenith angle of the sun; and $F_0$ is the band average of solar irradiance. The reflectance of foam and whitecaps can be estimated from wind speed by empirical relationship. Notably, the reflectance of foam and whitecaps is assumed to be substantially corrected by following aerosol correction (Gordon and Wang, 1994). The atmospheric correction in this study can be simplified as

$$
\rho_w = \frac{\rho_r - (\rho_r + \rho_a)}{t_0 t_i}
$$

where $\rho_r$ and $\rho_a$ are the reflectances resulting from Rayleigh and aerosol scattering, respectively. $\rho_r$ represents the interaction between the two, and multiple scattering ($\rho_a$) can also be included in the estimation. $t_0$ is the Sun-sea diffuse transmittance. $t_i$ is the sea-sensor diffuse transmittance. $\rho_w$ is the parameter of interest, the water-leaving radiance reflectance, which is defined as the multiplication product of $\pi$ and the water-leaving radiance divided by the above-water downwelling irradiance (Quinten and Ruddick, 2015).

The reflection values of the water quality stations were extracted from atmospherically corrected satellite imagery and used for analysis. Chosen pixel coordinates based on water quality sampling stations; however, it is unlikely to be ideal for the extraction of reflection values for several possible reasons, including errors between the coordinates and the actual sampling sites, noise caused by random debris on the surface of the reflected light from a nearby object pond area due to atmospheric scattering, and/or near the water sampling stations, which may be shallow, making reflections at the bottom of the problem. To obtain the reflection of the minimum error value, the search range was expanded to the entire pond area around the pixel designated to ensure that the position of the selected pixel in each image had the corresponding reflectance values. If two pixels have the same reflectance values, the pixel closest to the coordinates of the water quality sampling location, namely, the pixel in the center of the pond, was selected.

Multiple regression analysis

Multiple regression analysis was used in this study because it can generate results such as equations without using specialized software. Additionally, the relatively small sample size in this study does not warrant the use of ANNs. Correlation equations are derived to determine a quantitative relationship between the reflectance from each band of a given cell as the dependent variable and the concentration of the water component of interest, such as TP, in the cell as the independent variable. The SWIR band was used for the atmospheric correction, and consequently, band 6 and band 7 were not included in the regression. The other bands (band 1 to band 5) were used in subsequent calculations. In the regression analysis, the band ratios were used as independent variables because they are less likely to be affected by light conditions (Jensen, 2007). In the subsequent variable selection process, the initial independent
variables considered are tabulated in Table 3, where all band and band ratios are based on the reflectance.

The selection of predictor variables for TP was based on multiple regression analysis with forward selection. Variables are added to the regression one at a time in the forward selection, then starting with no predictor variables. The p-value threshold includes a predictor in the multiple regression equation if the p-value is lower than a "probability to enter". In addition, the predictor that will most improve the fit first is used, namely, in a "forward" approach. A default value of 0.05 in SPSS (IBM, 2019) was used for the "probability to enter". In the forward selection method, the variation inflation factor (VIF) should be further considered in the multiple regression analysis to avoid multicollinearity in the model. When a predicting variable is a linear combination of other prediction variables in the model, multicollinearity occurs. The direct consequence of multicollinearity is that the error variance is exaggerated. If the overfitted model is used with a new set of data, it may result in a low prediction power. The VIF is calculated as follows:

\[
VIF_j = \frac{1}{1-R_j^2}
\]  

where \( R_j \) is the multifactor coefficient of determination between the predictor variable of interest and others. A good rule of thumb to avoid severe multicollinearity is that all selected predictors should have VIF values of less than 10. (Chatterjee and Simonoff, 2013; Hocking, 2013). To include the VIF in the forward selection, the following procedure was proposed. First, make a forward selection. Second, we ensured that all the selected variables had a p value less than 0.05 and started to delete variables starting from the top of the p-value until all the variable p-values were as far as possible to below 0.05. Third, check the VIF of the remaining predictor variables, and if one of them has a value higher than 10, delete the variable with the highest VIF. Last, repeat step two; if all the variables of p-values are less than 0.05 and VIF values are less than 10, is to stop the process. Figure 2 illustrates the proposed procedure. As mentioned, when a variable was deleted from the model, the coefficients of variables, p-values, and VIFs were dynamically recalculated.

**Results And Discussion**

Multiple regression-derived equations were obtained using the band reflectance as a predictor variable to estimate the TP concentration as a water quality constituent. The best-fitting regression equation for TP chosen by forward selection is provided in Table 4. The results in Table 4 include p-values and VIF values for each of the response variables that is related to TP and the predictor variables that are on behalf of reflectance and 95% confidence intervals for the regression coefficients, standard error and the associated regression coefficients. Additionally, the names of the predictors are abbreviated; for instance, B1 is the reflectance of band 1 and B2/B1 is the ratio of reflectance in band 2 divided by that in band 1.

The resulting equation is
In which TP mg/L and B5 W/(m²μm). The information from band 5, the infrared frequency, has a significant important influence on the ratios of all the predictor variables. The coefficient of determination for the best-fitting equation was 0.67. Fuller et al. (2004) concluded that infrared bands can help predict chlorophyll-a concentrations, which are related to the trophic condition of the water and thus to the trophic level (i.e., TP), using band 7 and published in the USGS study. The observed versus multiple regression-predicted values of TP concentrations are plotted in Fig. 3.

To assess the value of water quality monitoring based on satellites, the water quality values in three irrigation ponds, as shown in Fig. 4 and Fig. 5, were predicted using Eq. 5. Figure 4 shows the spatiotemporal distribution of the TP concentrations in the ponds in the Taoyuan Tableland area. Notably, the locations on March 13, 2018, and May 16, 2018, are different from those on November 24, 2018, and April 17, 2019. Moreover, the locations on August 4, 2018, are different from those on the other days. The primary cause for these differences was cloud cover, which led to fewer ponds being visible in the imagery. Additionally, the areas of some ponds were too small to be detectable because the pixel size was 900 m². For the five image dates, Table 5 summarizes the descriptive statistics of the irrigation ponds, such as the mean, max, min, and standard deviation (SD) of the TP concentrations, and the number of ponds was counted on that day. Note that the descriptive statistics are based on the average of all the pixel values for each pond. The number of irrigation ponds varied due to cloud cover. Among the five images, the largest number of ponds observed on November 24, 2018, was 497 ponds, while the lowest number was 72, observed on May 16, 2018; the SD on May 16 was also higher than those of the other days, probably due to the small sample size. However, the TP concentration in Shetzu No. 1 was lower than that in Jao No. 32B on March 13, 2018, as shown in Fig. 5(a) and Fig. 5(b). The gray area depicts the extent of each pond. Both ponds had only a single pixel inside the corresponding grid cell because the floating photovoltaic power plant covered part of the pond surface. Thus, the spatial resolution of the observations became a limiting factor in satellite-derived water quality prediction.

Conclusions

This study selected Landsat 8 imagery to monitor the water quality of irrigation ponds in Taoyuan, Taiwan. A multiple regression model was derived using reflectance bands and collected data as predictor variables to estimate the TP concentration. The findings of this study are summarized below:

1. The coefficient of determination for the best-fitting equation is 0.67. The near-infrared band (band 5) has a strong influence on the predictive power of TP retrieval.
2. The retrieved Landsat 8 images for Shetzu No. 1 and Jao No. 32B have a limited number of pixels in the ponds because floating photovoltaic power plants cover a portion of the pond surfaces.
3. A multiple regression-derived equation was successfully used to map the TP concentrations, and the results validated the feasibility of using satellite remote sensing methods to monitor the water quality of the irrigation ponds.

Declarations

Acknowledgments

This work was financially supported by the Excellence Research Program - Core Consortiums (NTUCCP-109L901003, NTU-109L8836 of National Taiwan University (NTU), and NTU-109L891201), The Featured Areas Research Center Program of the NTU Research Center for Future Earth within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) of and the Ministry of Science and Technology of Taiwan (MOST108-2621-M-002-MY2).

Declarations

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent to Publish

Not applicable.

Authors' contributions

Conceptualization-Shang-Lien Lo; Methodology-Po-Wen Su, Shang-Lien Lo; Formal analysis and software-Po-Wen Su; Funding acquisition-Po-Wen Su and Shang-Lien Lo; Writing-original draft-Po-Wen Su. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

References
1. Barbini R, Colao F, Fantoni R, Palucci A, Borfecchia F, De Cecco L, Della Rocca AB, Martini S (1997) *Integration of passive and lidar remote sensed data in monitoring sea water quality*. Paper presented at the 7th International Symposium on Physical Measurements and Signature in Physical Measurements and Signatures in Remote Sensing, Courchevel, France

2. Bukata RP (2005) Satellite monitoring of inland and coastal water quality: retrospection, introspection, future directions. CRC Press, Boca Raton

3. Chou R-J, Wu C-T, Huang F-T (2017) Fostering multi-functional urban agriculture: Experiences from the champions in a revitalized farm pond community in Taoyuan, Taiwan. *Sustainability, 9*(11), 2097

4. Department of Land Administration (2002) Taiwan's geographic Aerial Map, 1:5,000 of scale in digital database forms. Taipei, Taiwan

5. Fang W-T, Chu H-J, Cheng B-Y (2009) Modeling waterbird diversity in irrigation ponds of Taoyuan, Taiwan using an artificial neural network approach. *Paddy Water Environ 7*(3):209–216

6. Fuller LM, Aichele SS, Minnerick RJ (2004) Predicting water quality by relating secchi-disk transparency and chlorophyll-measurements to satellite imagery for Michigan inland lakes, August 2002. U.S. Geological Survey, Reston

7. Gordon HR, Wang M (1994) Influence of oceanic whitecaps on atmospheric correction of ocean-color sensors. *Appl Opt 33*(33):7754–7763

8. Harmel R, King K, Haggard B, Wren D, Sheridan J (2006) Practical guidance for discharge and water quality data collection on small watersheds. *Transactions of the ASABE 49*(4):937–948

9. IBM (2019) IBM SPSS Statistics 26 Documentation. IBM, Endicott

10. Jensen J (2007) Remote sensing of the environment: an earth resource perspective, 2nd edn. Pearson Prentice Hall, Upper Saddle River

11. Kannel PR, Lee S, Kanel SR, Khan SP, Lee Y-S (2007) Spatial–temporal variation and comparative assessment of water qualities of urban river system: a case study of the river Bagmati (Nepal). *Environ Monit Assess 129*(1–3):433–459

12. Kishino M, Tanaka A, Ishizaka J (2005) Retrieval of Chlorophyll a, suspended solids, and colored dissolved organic matter in Tokyo Bay using ASTER data. *Remote Sens Environ 99*(1–2):66–74

13. Kloiber SM, Brezonik PL, Olmanson LG, Bauer ME (2002) A procedure for regional lake water clarity assessment using Landsat multispectral data. *Remote Sens Environ 82*(1):38–47

14. Liu Y, Islam MA, Gao J (2003) Quantification of shallow water quality parameters by means of remote sensing. *Prog Phys Geogr 27*(1):24–43

15. Markham B, Barsi J, Kvaran G, Ong L, Kaita E, Biggar S, Czapla-Myers J, Mishra N, Helder D (2014) Landsat-8 operational land imager radiometric calibration and stability. *Remote Sensing 6*(12):12275–12308

16. Pahlevan N, Garrett AJ, Gerace AD, Schott JR (2012) Integrating Landsat-7 imagery with physics-based models for quantitative mapping of coastal waters near river discharges. *Photogrammetric Engineering Remote Sensing 78*(11):1163–1174
17. Schott JR, Barsi JA, Nordgren BL, Raqueño NG, de Alwis D (2001) Calibration of Landsat thermal data and application to water resource studies. Remote Sensing of Environment 78(1):108–117

18. Song L-Y, Yadav R, Liang H-C (2018) Research on Eco-friendly Solar Energy Generation in Taoyuan Pond. Paper presented at the 2018 IEEE International Conference on Advanced Manufacturing (ICAM)

19. Vanhellemont Q (2017) ACOLITE Manual version 20170718.0. Royal Belgian Institute of Natural Sciences, Brussels

20. Vanhellemont Q, Ruddick K (2015) Advantages of high quality SWIR bands for ocean colour processing: Examples from Landsat-8. Remote Sens Environ 161:89–106

Tables

**Table 1** Days of the satellite images and water quality samples.

| Water Quality Sampling Date | Satellite Image Date |
|-----------------------------|----------------------|
| February 23, 2018           | March 13, 2018       |
| May 16, 2018                | May 16, 2018         |
| August 14, 2018             | August 4, 2018       |
| November 9, 2018            | November 24, 2018    |
| May 14, 2019                | April 17, 2019       |

**Table 2** Band attributes of Landsat 8 OLI.

| Spectral Band | Wavelength (μm) | Resolution (m) |
|---------------|-----------------|----------------|
| Band 1        | 0.43-0.45       | 30             |
| Band 2        | 0.45-0.51       | 30             |
| Band 3        | 0.53-0.59       | 30             |
| Band 4        | 0.64-0.67       | 30             |
| Band 5        | 0.85-0.88       | 30             |
| Band 6        | 1.57-1.65       | 30             |
| Band 7        | 2.11-2.29       | 30             |
| Band 8        | 0.50-0.68       | 15             |
| Band 9        | 1.36-1.38       | 30             |

**Table 3** Initial band reflectance values and band reflectance ratios considered during the variable selection process (“B” is the abbreviation of “band”).

| Water constituent | Number of valid observations | Initial prediction variables before the p-threshold test |
|-------------------|------------------------------|--------------------------------------------------------|
| TP                | 6                            | B1, B2, B3, B4, B5, B2/B1, B3/B1, B4/B1, B5/B1, B3/B2, B4/B2, B5/B2, B4/B3, B5/B3, B5/B4 |

**Table 4** Best-fitting multiple regression model for TP using forward selection based on p-values and VIFs.
| R² | Number of obs. | Predictor | Coefficient of the predictor | VIF |
|----|----------------|-----------|------------------------------|-----|
|    |                |           | Value | Std. Error | Lower 95% | Upper 95% | p   |
| TP | 0.67           | 6         | Intercept | -0.023 | 0.041 | -0.138 | 0.091 | 0.600 | n/a |
|    |                | B5        | 5.020 | 1.763 | 0.124 | 9.915 | 0.047 | 1    |

Table 5 Descriptive statistics for five images.

| Date               | Number of ponds | TP concentration (mg/L) |
|--------------------|-----------------|-------------------------|
|                    |                 | Mean | Max  | Min  | SD    |
| March 13, 2018     | 194             | 0.184 | 0.273 | 0.082 | 0.026 |
| May 16, 2018       | 72              | 0.079 | 0.304 | 0.004 | 0.053 |
| August 4, 2018     | 112             | 0.053 | 0.206 | 0.002 | 0.036 |
| November 24, 2018  | 547             | 0.051 | 0.269 | 0.002 | 0.035 |
| April 17, 2019     | 497             | 0.048 | 0.252 | 0.001 | 0.035 |