Monitoring of regional lake water clarity using Landsat imagery
Matias Bonansea, Raquel Bazán, Claudia Ledesma, Claudia Rodriguez and Lucio Pinotti

ABSTRACT

The application of remote sensing technology to water quality monitoring has special significance for lake management at regional scale. Water clarity expressed in terms of Secchi disk transparency (SDT) is a highly useful indicator of trophic status and ecosystem health. In this study, we related Landsat TM and ETM+ data with ground observations to develop a model for the estimation of SDT which can be used as a standardized procedure for regional-scale lake clarity assessment in the central region of Argentina. Samples were taken from two reservoirs of the region. Pearson correlation coefficients and step-wise multiple regression analysis were used to evaluate correlation between Landsat bands and measured SDT. Results suggested that Landsat band 3 plus the ratio 1/3 was a consistent and reliable predictor of SDT ($R^2 = 0.80$). The algorithm was validated ($R^2 = 0.81$) and applied to the November 10, 2010 ETM+ image obtaining a map that characterized water clarity of reservoirs within the study area. The procedure presented here could become a low cost measurement tool for water management authorities and decision-makers, obtaining simpler and practical results for regional water clarity monitoring.

Key words | algorithm, Landsat, remote sensing, reservoir, Secchi disk transparency, water clarity

INTRODUCTION

Decision-makers are demanding new tools for regional monitoring and assessment of water quality. The conventional measurements of regional assessment are logistically challenging and expensive to perform regularly due to cost, lake accessibility and the number of water bodies requiring repeated sampling (McCullough et al. 2012b). As a result, sample sizes must be limited and usually cannot encompass each type of water body present in a region; therefore, the status of the water system at a regional scale can be difficult to represent (Zhao et al. 2011). According to McCullough et al. (2012a), these restrictions lead to field assessments concentrated in developed, easily accessible areas, which create spatially irregular, non-random samples. Many lakes are rarely or never monitored, so an accurate assessment of their status and change over time cannot be made. Satellite remote sensing has been shown to be a powerful supportive tool for regional water quality assessment, reducing costs and allowing monitoring to occur simultaneously across an extensive area (Trivero et al. 2013; Larsen et al. 2013; Doña et al. 2014).

Among several satellite systems that have been used for water quality monitoring, the Landsat system, which provides an unparalleled record of the status and dynamics of the Earth’s surface since 1972, is particularly

doi: 10.2166/nh.2014.211
useful for assessment of inland lakes (Kloiber et al. 2002b; Matthews 2010; Wulder et al. 2012; Chao Rodríguez et al. 2014). Most techniques for remote sensing of water quality construct reliable empirical relationships between Landsat data and ground observations of water quality parameters, including chlorophyll and phycocyanin concentrations (Vincent et al. 2004; Tebbs et al. 2013), water clarity expressed in terms of Secchi disk transparency (SDT) (Domínguez Gómez et al. 2009; Zhao et al. 2011; McCullough et al. 2012a), total suspended sediments (Kulkarni 2011; Bonansea & Fernandez 2015), among others. In this study, we focus on SDT estimation due to its simplicity and relatively low cost. Besides, this parameter, which is widely used and a common metric of lake water quality, has strong ecological and economic implications, being a highly useful indicator of trophic status and ecosystem health (Sriwongsitanon et al. 2011; Zhao et al. 2011; McCullough et al. 2012b; Chao Rodríguez et al. 2014). According to Domínguez Gómez et al. (2009), the assessment of water clarity has a crucial impact on water quality monitoring because it shows, in a global way, all the components that can be found in water and the many interactions existing among them.

Most studies related to water clarity estimation by Landsat imagery have focused on generating empirical models for the lake or reservoir where samples were taken (Domínguez Gómez et al. 2009; Giardino et al. 2010; Guan et al. 2011). However, there has been increasing focus on regional-scale assessment of water quality and few monitoring programs exist for this purpose. Pulliainen et al. (2001) suggest that the estimation of water quality from remote sensing data for numerous lakes could be achieved using ground observation data for only a few representative lakes from the region. Kloiber et al. (2002b) and Olmanson et al. (2008) described a practical and efficient procedure for Landsat imagery for routine, regional-scale assessments of lakes for water clarity, and Kloiber et al. (2002a) used this approach to measure spatial patterns and temporal trends in a large number of lakes. McCullough et al. (2012a) have shown that Landsat TM can be used to predict regional water clarity in Maine lakes located in the northeastern United States, and those predictions are more accurate when average depth and watershed wetland area are included in models. Although most of the studies on regional lake water clarity estimation by remote sensing were carried out for the northern hemisphere, little has been done to develop appropriate regional assessment of water clarity in the southern hemisphere.

The objective of this paper was to develop an algorithm to estimate water clarity which can be used as a standardized procedure for regional-scale lake clarity assessment in the central region of Argentina. Thus, we were able to obtain a single standardized method with constant coefficient values that could be used by water management authorities and decision-makers to achieve information for lakes not sampled, allowing an easier comparison of water clarity from different lakes at a regional scale.

**METHODOLOGY**

**Study area**

The western region of Córdoba province, located in the central region of Argentina, is characterized by a mountainous system called Sierras Pampeanas which encompasses approximately a 500 km long and 150 km wide area. This area presents nine moderately eutrophic reservoirs greater than 0.3 km² (Figure 1) which were built between 1950 and 1950 for multiple purposes such as water supply, power generation, flood control, irrigation, tourism and recreational activities (Bazán et al. 2005; Mancini et al. 2010; Ledesma et al. 2013).

As part of a monitoring program, since the 1990s several physical, chemical and biological properties of two multipurpose reservoirs of the area (Río Tercero and Los Molinos reservoirs) have been surveyed. Río Tercero reservoir (32° 11’ S, 64° 23’ W) which is the largest artificial reservoir in the province, has a surface area of 46 km², a volume of 733 hm³ and maximum and mean depths of 46.5 and 12.2 m, respectively. In 1986, a nuclear power plant (CNE: 600 MWA) was installed. Water for cooling the nuclear reactor is taken from the middle section of the reservoir and is returned to the western basin by a 5 km long open-sky channel (Bonansea et al. 2014). Los Molinos reservoir (31° 49’ S, 64° 32’ W), which is used to
supply drinking water to Córdoba city (with 1.4 million inhabitants), has an area of 21.1 km², an average depth of 16.3 m and a maximum volume of 399 hm³ (Bazán et al. 2018).

Water clarity was estimated in the field by measuring SDT using a standard 20 cm diameter Secchi disk at nine sampling sites in Río Tercero and five sites in Los Molinos reservoir (Figure 1). Coordinates of sample sites were recorded using a Global Positioning System (GPS) device.

**Satellite data**

We used images from Landsat TM and Landsat ETM+ (Path: 229; Row: 82) downloaded from the USGS Global Visualization Viewer (http://glovis.usgs.gov). The TM sensor is equipped with multi-spectral scanning equipment, which operates on seven spectral bands located between the visible and infrared regions of the spectrum. The spatial resolution is 30 m for the visible through middle infrared channels and 120 m for the thermal infrared band (Loveland & Dwyer 2005). The ETM+ sensor has a similar suite of bands as TM, but with a 60 m thermal band and an additional 15 m panchromatic band. Both sensors present a revisit time of 16 days and a radiometric resolution of 256 digital numbers.

The criteria for image selection were: existing in situ data of both reservoirs in ±4 days to the satellite passes (time window) obtaining reasonable results for empirical relationships between SDT and Landsat imagery; no heavy rainfall prior to the image data to minimize the effects of changes in water surfaces that disturb the estimates; 0% haze or cloud cover when possible. To detect haze and cloud cover, which affect spectral-radiometric responses and cause erroneous results, an RGB band combination (1,6,6) was used (Olmanson et al. 2008). The selected criteria are in agreement with different authors (Kloiber et al. 2002b; Sriwongsitanon et al. 2011;...
Thus, from the pool of suitable images, we selected one Landsat TM image and two ETM+ images (Table 1).

**Image pre-processing**

The electromagnetic radiation signals collected by satellites in the solar spectrum are modified by scattering and absorption by gases and aerosols while traveling through the atmosphere from the Earth’s surface to the sensor (Song et al. 2001). Atmospheric corrections to satellite data are therefore important for correcting these effects, so that information from multitemporal data set with variable aerosol loading can be sensibly compared. Using the Second Simulation of the Satellite Signal in the Solar Spectrum (6S) (Vermote et al. 1997), atmospheric correction was carried out.

The importance of applying the 6S model to improve the estimates of water lake clarity was described in Sriwongsitanon et al. (2011). These authors suggest that the 6S model can remove the additive effects provided by atmospheric rayleigh and aerosol scattering which influence the visible Landsat bands (band 1–3). On the other hand, the corrected reflectance values of the near infrared and infrared bands (band 4–7) tend to be higher than the uncorrected reflectance values. This is because the near infrared and middle infrared wavelengths are affected by atmospheric absorption while the influence of air molecules and aerosol particle scattering are negligible in these ranges. Since the 6S model can remove these effects, reflectance values within these bands were then increased (Sharma et al. 2009; Sriwongsitanon et al. 2011; Homem Antunes et al. 2012).

Geometric correction was applied to each scene, resulting in a root mean square error (RMSE) of positional accuracy of less than 0.5 pixel, guaranteeing a precise geometric match between images. Since May 2003, ETM+ images have a permanent failure known as Scan Line Corrector (SLC): off, characterized by wedge-shaped gaps (Chen et al. 2011). Using a methodology adapted from the SLC Gap-Filled Products, Phase One Methodology article (USGS 2004), SLC failure was corrected predicting the best closest value of the missing pixels. To delineate the lake surface masks, producing ‘water-only’ images and isolating anomalously pixels that do not belong to the reservoirs, the normalized difference water index (NDWI) algorithm proposed by McFeeters (1996) was applied. According to Ji et al. (2009) and Alcântara et al. (2010), the NDWI can be used successfully in delineating water bodies and monitoring the water area changes.

**Algorithm development**

As the locations of sampling points were georeferenced, it was possible to compare matchups between field data and corresponding Landsat reflectance values. To determine which spectral band or band ratio was the best predictor of SDT, Pearson correlation coefficients and backward step-wise multiple regression analysis were carried out between in situ SDT (dependent variable) versus atmospherically corrected reflectance values of Landsat bands or band ratios (independent variables). Applying the Pearson correlation analysis, we assume that a high level of correlation between variables is implied by a correlation coefficient (r) greater than 0.5 in absolute terms (Gupta 1999). The backward step-wise multiple regression analysis was performed using the thresholds for factor removal with a significance level of p-value more than 0.05. If the p-value is less than the threshold, it means that the null hypothesis is rejected and the regression relationship is then reliable to be used for prediction (Sriwongsitanon et al. 2011). Thus, we could identify the spectral band or band ratio most correlated with in situ SDT, which were used to generate a model to estimate water clarity for all lakes within the region. In this case, the multiple linear regression model used was

<table>
<thead>
<tr>
<th>Acquisition image date</th>
<th>Landsat sensor</th>
<th>Sampling date</th>
<th>Time window (days)</th>
<th>Sampling date</th>
<th>Time window (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>09-28-2006</td>
<td>ETM +</td>
<td>09-24-2006</td>
<td>–4</td>
<td>10-01-2006</td>
<td>3</td>
</tr>
<tr>
<td>12-09-2006</td>
<td>TM</td>
<td>12-09-2006</td>
<td>0</td>
<td>12-11-2006</td>
<td>2</td>
</tr>
<tr>
<td>11-10-2010</td>
<td>ETM +</td>
<td>11-10-2010</td>
<td>0</td>
<td>11-10-2010</td>
<td>0</td>
</tr>
</tbody>
</table>
defined as

\[ Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \cdots + \beta_v X_v + \epsilon \]  

(1)

where \( Y_i \) refers to the response of the variable SDT, \( X_n \) are the explanatory variables of each Landsat spectral bands, \( \beta_n \) are the regression coefficients, and \( \epsilon \) is the random error.

Simple regression analysis was made to evaluate the correlation between estimated versus observed SDT data. The RMSE of predicted SDT, which gives an estimate of the error associated with the estimations (Matthews et al. 2010), was calculated according to

\[ \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}} \]  

(2)

where \( X_i \) and \( X \) are the in situ and satellite-derived SDT and \( n \) is the sampling size.

Finally, as a demonstration of the real potential of remote sensing, the validated algorithm was applied to the pre-processed November 10, 2010 ETM+ image, obtaining the spatial distribution of simulated errors, calculated as the difference between simulated and observed SDT data, and a map that characterizes water clarity of reservoirs in the central region of Argentina.

RESULTS AND DISCUSSION

Estimation of water clarity

Remote sensing of water quality parameters is dependent upon how parameter variations alter the optical properties of the water column (Pavelsky & Smith 2009). Figure 2 shows the results of Pearson correlation coefficient and step-wise regression analyses between SDT versus Landsat spectral bands. According to Brezonik et al. (2005), suspended particles cause and increase in the measured response for Landsat bands 1–4. Landsat band 1, which can be used to measure the irradiance attenuation due to the absorption of aquatic humus and phytoplankton pigment concentration (Giardino et al. 2001), showed a low association with measured SDT \((r = -0.39)\). Landsat band 2 is centred on an algal reflectance peak (Brezonik et al. 2005; Domínguez Gómez et al. 2009). This spectral band, which was widely used for estimating chlorophyll-a concentration (Domínguez Gómez et al. 2009; Kulkarni 2011), showed a low relation with measured SDT \((r = -0.39)\). A low negative association was found between SDT and band 4 \((r = -0.47)\); this could be explained because absorbance by water increases sharply in this band (Brezonik et al. 2005). There was no association between measured SDT and Landsat bands 5 and 7 \((r = -0.02 \text{ and } -0.10\), respectively). Srivongsitanon et al. (2011) suggest that in the infrared regions (band 5 and 7), water increasingly absorbs the light making it darker so these bands are useful for vegetation and soil moisture studies and for discriminating between rock and mineral types. Therefore, we have not analyzed the band ratios of these bands. The thermal infrared band of TM and ETM+ sensors (band 6) was not used in the analysis because this band, which is based on the reflective properties of the Earth’s surface in the short-wave part of the electromagnetic spectrum, is used to estimate surface temperature (Giardino et al. 2001; Chao Rodríguez et al. 2014). Our results demonstrated that Landsat band 3 \((\rho_3)\) and band ratio 1/3 \((\rho_1/\rho_3)\) can be used to investigate the most suitable relationships for SDT monitoring as evidenced by high Pearson correlation coefficients \((r = -0.78 \text{ and } 0.79\), respectively). These results were confirmed in the step-wise multiple regression analysis where only band 3 and ratio band 1/3 were retained \((R^2 = 0.80)\). According to Matthews (2010), the negative correlation with band 3
may be explained by the direct positive correlation between reflectance in the red and gross particulate load inducing particulate scattering. Therefore, as SDT decreases, brightness in the red usually increases.

Thus, the estimated response between in situ SDT and the atmospherically corrected reflectance values in $\rho_3$ and $\rho_1/\rho_3$, was finally formulated as ($R^2 = 0.80$)

$$SDT = 3.22 - 1.66\rho_3 + 0.64\rho_1/\rho_3$$

(3)

The 95% confidence intervals for the parameters of the model (Equation (1)) were calculated as $2.35 < \beta_0 < 4.09$; $-2.32 < \beta_1 < -1.00$; and $0.41 < \beta_2 < 0.88$.

Values of estimated and observed SDT were correlated applying a simple regression model. The good fit between observed and estimated SDT indicated the high predictive capacity of this model ($R^2 = 0.80$). The error associated with the estimations (RMSE = 0.64 m) was also reasonable and lower than the RMSE in SDT measured in McCullough et al. (2012a, 2012b). Figure 3 also confirms the robustness of this algorithm as giving a good agreement between the gradient and intercept of the regression line. Therefore, the methodology used was considered to be adequate to study water clarity assessment in different water bodies of the region.

According to Matthews (2010), there are a large number of studies using Landsat to retrieve SDT, and most of these use linear regressions of single bands or band ratios. Different studies suggest that SDT can be estimated from different combinations of Landsat bands 1 to 4 (Doña et al. 2014). Some studies use Landsat band 2 (green band) or 4 (NIR band) to estimate SDT (Lathrop & Lillesand 1986, 1989; Doña et al. 2014), although there are few recent examples of this (Matthews 2010). Domínguez Gómez et al. (2009), studying the trophic status of lakes located to the south of Madrid, Spain, found that SDT, which is affected by phytoplankton and total suspended solids concentration, could be associated with Landsat band 2, which shows the highest light penetration. However, in our study, Landsat band 3 plus the ratio 1/3 provided strong predictive relationship with SDT in reservoirs of Córdoba province. Several investigators had success with similar relationship. The same band combination was used by Lavery et al. (1995) studying an estuarine system in western Australia. Hellweger et al. (2004) found that TM band 3 provided a strong relationship to SDT. McCollough et al. (2012a) used TM bands 1 and 3 to predict SDT for Maine lakes, United States. According to Matthews (2010), the ratio between TM bands 1 and 3 is particularly common to estimate lake water clarity. Lathrop (1992) and Cox et al. (1998) suggest that ratio 1/3 is a strong predictor of SDT. Kloiber et al. (2002b) and Brezonik et al. (2005) used Landsat band 1 plus ratio 1/3 to predict SDT with high accuracy. Similar results were found by Olmanson et al. (2008) studying a series of lakes in Minnesota, United States and Zhao et al. (2011) in Taihu lake, China.

Map generation

The analysis of the spatial distribution of simulation errors of the November 10, 2010 ETM+ image indicated that the central region of the reservoirs showed a lower difference between simulated and observed data (Figure 4). Both reservoirs showed that higher simulation errors, which were located near the shores and tributaries, could be related with the effect of the bottom or with tributaries inflow which generate an important change in SDT (Bonansea & Fernandez 2013). Although the method lost accuracy, the trend curve continues to be coherent.

Although the validated algorithm was based on ground observations from only two reservoirs of the region, we used it to estimate SDT for all lakes within the study area. This is in agreement with McCullough et al. (2012a) who
demonstrated that remote sensing is useful in regions containing a large number of lakes that are cost prohibitive to monitor regularly using traditional field methods. Thus, we obtained a complete regional spatial sampling of water clarity, allowing the mapping and analysis of spatial patterns. Figure 5 shows the spatial distribution of SDT in Córdoba reservoirs applying the validated algorithm to the pre-processed November 10, 2010 ETM+ image. Satellite estimated SDT ranged from $<0.5$ to about 6.7 m, with a mean value of 3.5 m. Lower SDT values were registered in San Roque reservoir, coinciding with Amé et al. (2003) and Galanti et al. (2015), who suggest that this reservoir is classified as eutrophic to hypereutrophic with elevated concentrations of nutrients. Lower SDT were also registered in Arroyo Corto reservoir whose waters are pumped to Cerro Pelado reservoir and reused to generate energy by the Cerro Pelado Hydroelectric Complex in times of low rainfall, as the case of this date. Dividing the reservoirs into classes based on 1.5 m SDT intervals, we found that the most common clarity class is 3.0 to 4.5 m. Mean water clarity remained stable between the reservoirs, with the exception of Arroyo Corto.

Limitations

Landsat sensors are a powerful tool that can provide systematic and periodic information of water clarity in reservoirs. However, there are limitations to monitoring water quality with Landsat imagery. Over the past decade, TM and ETM+ imagery availability decreased over time due to different problems or anomalies (Wulder et al. 2011; Marx & Loboda 2013). The deteriorated image quality resulting from SLC failure, which was mentioned before, has become a major obstacle for Landsat 7 data applications.
Since 2005, Landsat 5 has had problems with its solar array drive which has affected data availability (Wulder et al., 2011) and in mid-2013 this satellite was officially decommissioned. However, Landsat 5 is the longest-operating Earth-observing satellite mission in history, transmitting over 2.5 million images of land surface conditions around the world and resulting in a unique, long-term, systematic collection of moderate resolution imagery (Wulder et al., 2015). As we have included a TM image, we could perform a retrospective analysis of water clarity back to the early 1980s, since Landsat 5 was launched, and surface data were obtained by TM sensor.

The scope of our study may be expanded with the inclusion of the new Landsat 8 LDCM which was launched on January 2013 and is the follow-on mission to Landsat 7, presenting a higher imaging capacity than previous Landsat satellites (Loveland & Dwyer, 2012; Wulder et al., 2012). Although there are no other missions analogous to Landsat with global observation capabilities or accumulated global archives, Wulder et al. (2011) suggest that several programs and sensors are identified as having the potential to emulate Landsat sensors. Thus, other satellite remote sensors, such as SPOT, ASTER, MODIS, MERIS, may be useful alternatives for lake water quality (McCullough et al., 2012b; Chawira et al., 2015). However, a careful comparison of remote sensing reflectance data between sensors would be required beforehand and specific user needs should guide the selection of alternate data source.

Remote sensing provides suitable information concerning water quality and aquatic systems management. In this study, we demonstrated that water clarity of Córdoba reservoirs can be estimated by Landsat imagery. Pearson correlation coefficients and step-wise multiple regression analyses were used to investigate the relationship between SDT versus Landsat bands and band ratios. Band 3 and the ratio 1/3 proved to be consistent predictors of water clarity. The obtained algorithm was used as a standardized procedure for regional-scale lake clarity assessment in the central region of Argentina.

Rather than using regressions equations where the independent variables and coefficients are different for each Landsat image, we examined the feasibility of using a consistent water clarity equation form to relate ground observation and satellite data. Use of a consistent equation form is preferable because it allows for easier comparison of the results from different images. Thus, the procedure presented here could become an independent, low additional training and low cost measurement tool for water management authorities and decision-makers, obtaining simpler and practical results for regional water clarity monitoring. However, the implementation and continuation of field-based reservoir water quality monitoring in Córdoba reservoirs is essential for better calibration and validation of future remote clarity estimation models. Finally, the inclusion of the new LDCM or other potential satellite sensors (e.g. SPOT, ASTER, MODIS, and MERIS) could be useful to extend our study.

**CONCLUSIONS**

**ACKNOWLEDGEMENTS**

The authors thank the editor and reviewers for their helpful comments on this manuscript. This work was supported by
SECyT-UNRC (Secretaría de Ciencia y Técnica, UNRC) and SECyT-UNC. Additional financial support was provided by CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas).

REFERENCES


First received 13 June 2014; accepted in revised form 15 August 2014. Available online 20 September 2014