Generation of the bathymetry of a eutrophic shallow lake using WorldView-2 imagery
Onur Yuzugullu and Aysegul Aksoy

ABSTRACT

In this study, water depth distribution (bathymetric map) in a eutrophic shallow lake was determined using a WorldView-2 multispectral satellite image. Lake Eymir in Ankara (Turkey) was the study site. In order to generate the bathymetric map of the lake, image and data processing, and modelling were applied. First, the bands that would be used in depth prediction models were determined through statistical and multicollinearity analyses. Then, data screening was performed based on the standard deviation of standardized residuals (SD_SR) of depth values determined through preliminary linear regression models. This analysis indicated the sampling points utilized in depth modelling. Finally, linear and non-linear regression models were developed to predict the depths in Lake Eymir based on remotely sensed data. The non-linear regression model performed slightly better compared to the linear one in predicting the depths in Lake Eymir. Coefficients of determination ($R^2$) up to 0.90 were achieved. In general, the bathymetric map was in agreement with observations except at re-suspension areas. Yet, regression models were successful in defining the shallow depths at shore, as well as at the inlet and outlet of the lake. Moreover, deeper locations were successfully identified.

Key words | bathymetry, remote sensing, shallow lake, WorldView-2

INTRODUCTION

Water depth is important for several physical and biological processes in a lake (Leira & Cantonati 2008). Together with water volume, water depth impacts natural assimilation capacity, pollution dilution factor, water temperature and retention time. Light penetration and growth of algal species (especially attached algae) may depend on the depth of water. Furthermore, water depth influences mixing of water layers, sedimentation of solids and re-suspension of bottom sediments. Therefore, obtaining the spatial distribution of water depths or bathymetric information may be critical in assessing the impact of pollutants on lake water quality.

Sonar/radar systems have been frequently used in derivation of the bathymetric maps of lakes (Tureli & Norman 1992; Morgan et al. 2003). However, these systems may require extensive fieldwork and financial means, especially for large water bodies. In order to ease these difficulties, remotely sensed images (hyperspectral, multispectral, aerial and radar) can be used in obtaining bathymetric information. In the literature, such applications have mainly focussed on coastal waters and estuaries (Philpot 1989; Greidanus et al. 1997; Robbins 1997; Hennings 1998; Sandidge & Holyer 1998; Roberts 1999; Calkoen et al. 2001; Lafon et al. 2002; Dierssen et al. 2003; Stumpf et al. 2003; Jordan & Fonstad 2005; Mobley et al. 2005; Lyzenga et al. 2006; McIntyre et al. 2006; Bachmann et al. 2008, 2009; Kao et al. 2009; Lee 2010; Marchisio et al. 2010). However, the use of multispectral images in determination of the bathymetry of lakes is not common. This is due to the fact that complex relationships dominate between radiance and lake characteristics in lakes compared to Case 1 waters, as well as coastal waters and estuaries. Complexities arise from high chlorophyll-a concentrations, suspended solids, organic matters and bottom reflection in most lakes. The existence of few sensors on most multispectral satellites may result in insufficiency in distinguishing between the impacts of these factors.
on radiance values. However, the launch of new satellites, such as WorldView-2, may provide new means and additional sensors that may aid in depth determination in lakes as well.

The WorldView-2 satellite was launched in the fourth quarter of 2009. It has eight spectral bands covering the electromagnetic spectrum range of 400–1,040 nm (Table 1) (Digital Globe 2010). With its 0.46 m panchromatic and 1.84 m multispectral resolution, studies which require high spatial resolution can be conducted (Lee et al. 2011). The satellite has a radiometric resolution of 11-bits and a temporal resolution of 3.7 days at 20° or less. It allows images to be captured in an area of 65.6 km × 110 km at the nadir. Its coastal blue band that senses the 400–450 nm of the spectrum is characterized by its relatively shorter wavelength and higher energy. It can penetrate to deeper parts of water bodies. It has been reported that depths down to 30 m can be identified by coastal blue and blue bands (Digital Globe 2010).

WorldView-2 imagery has been used in recent studies for bathymetry determinations in coastal waters and estuaries (Glass et al. 2010; Marchisio et al. 2010; Lee et al. 2011; McCarthy et al. 2011; Parthish et al. 2011). Lee et al. (2011) reported that green and yellow bands were more effective in depth determination in the range of 2.5–20 m in coastal waters. Marchisio et al. (2010) showed the efficacy of coastal blue and blue bands in revealing depths up to 7 m. To our knowledge, there is no study on bathymetry generation for shallow eutrophic lakes using the WorldView-2 imagery.

In this study, a WorldView-2 image was used to determine the bathymetry and, therefore, the spatial distribution of water depths in eutrophic Lake Eymir in Ankara, Turkey. The relationships between measured depths and radiance values at different bands were investigated.

<table>
<thead>
<tr>
<th>Band</th>
<th>Wavelength range (nm)</th>
<th>Band</th>
<th>Wavelength range (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coastal Blue (Band 1)</td>
<td>400–450</td>
<td>Red (Band 5)</td>
<td>630–690</td>
</tr>
<tr>
<td>Blue (Band 2)</td>
<td>450–510</td>
<td>Red Edge (Band 6)</td>
<td>705–745</td>
</tr>
<tr>
<td>Green (Band 3)</td>
<td>510–580</td>
<td>NIR-1 (Band 7)</td>
<td>770–895</td>
</tr>
<tr>
<td>Yellow (Band 4)</td>
<td>585–625</td>
<td>NIR-2 (Band 8)</td>
<td>860–1,040</td>
</tr>
</tbody>
</table>

Linear and non-linear regression models were developed to derive the spatial distribution of depths. The depths obtained with these models were compared to the actual bathymetric map of the lake.

**MATERIALS AND METHODS**

**Study area**

Lake Eymir is a shallow natural lake located at 39.28 N and 32.30 E. It is located 20 km south of Ankara (Figure 1) at an altitude of 969 m (Beklioglu et al. 2003). The surface area of the lake is around 1.25 km². It has a shoreline of 11 km and a catchment area of 971 km². In 1990, the area surrounding the lake and 245 km² of its catchment area was declared as a ‘Special Environmental Protection Area’ by Decree of the Cabinet of Ministers due to its ecological significance (Yuzugullu 2011).

The average water depth in the lake changes depending on the balance between inflows and outflow. Lake Eymir is mainly fed by Lake Mogan in the south (98% of the total inflow), the Kislakci Stream in the east and groundwater sources. The excess water of the lake drains into Imrahor Creek in the east (Yenilmez et al. 2011). The average water depth is 4 m. Annual water level fluctuations in the lake vary by 0.5–1.0 m, depending on the net inflow and evaporation (Yagbasan & Yazicigil 2009). In April 2011, the average depth in the lake reached 4.5 m (Yuzugullu 2011).

The lake has been suffering from the effects of eutrophication. It has been turbid and rich in algal species for a long time. In studies performed in different time periods, eutrophic conditions were reported (Diker 1992; Tan 2002; Ozen 2006). It was shown that water balance, and therefore the depth of water, had a significant impact on the water quality of the lake (Beklioglu et al. 2003). Therefore, bathymetry and water depths provide substantial information in the assessment of water quality changes in the lake.

**METHODOLOGY**

The methodology followed in the development of depth (or bathymetry) models based on remotely sensed data is...
depicted in Figure 2. Following the acquisition of the image on 28 July 2010, a field work was realized on 2 August 2010. Depth measurements were conducted at 59 points (depicted in Figure 1). In the time gap between the image and field work dates, there was no precipitation or significant change in temperature or other conditions which would alter water depths. As a result, it was assumed that the depths and water quality parameters were representative of the conditions on the date the image was taken. Sampling locations for ground truth data were selected arbitrarily to cover the whole lake area. The geographical coordinates of the points were determined using a Garmin GPS receiver with ±1.5 m positional accuracy on average.

Image processing was conducted using ENVI 4.7. The image had geographic projection and ED 50 Datum. The lake area was cropped and isolated. Therefore, only the lake area was taken as the region of interest. The dark pixel subtraction method was used in order to eliminate atmospheric effects in the ortho-rectified satellite image (Chavez 1975). In order to obtain the radiance values, first, image histograms were generated for the corresponding spectral bands. Then, zero values in the histograms were removed. Finally, the minimum and the maximum values in the histograms were determined to aid in the conversion.

Figure 1 | Locations of Lake Eymir and sampling points.

Figure 2 | Flowchart of the methodology.
of digital numbers to radiance values using the method provided by Beisi et al. (2008). Histogram values were matched to digital numbers in the range of 0–255. By generating band-specific linear equations, digital numbers were converted to radiance values.

In order to test the suitability of data in regression model development and to improve the model prediction performance, a screening procedure was applied to select the independent variables of depth models. The initial stage of the procedure was to remove the sampling points at locations with high turbidity. This was applied to minimize the negative impact of re-suspended sediments in depth determination. Since Lake Eymir is a shallow lake, local re-suspension can occur due to various factors such as groundwater inflow, wind effect, turbulence due to velocity variations as a result of cross-sectional and flow direction changes. As depicted in Figure 1, the shape of Lake Eymir makes it prone to these impacts. On the sampling date, the lake was mostly clear with an average total suspended solids concentration of 1.92 mg/L and an average chlorophyll-a concentration of 3.49 μg/L, respectively (Yuzugullu 2011). However, at some locations turbidity was observed due to re-suspension of bottom sediments. These locations were identified on the image by locating the zones that exhibit high radiance due to suspended solids. The sampling points were placed over the image and the ones that were over the re-suspension areas were removed from the data set. As a result, 11 sampling points were removed from further analysis. These points can be seen in Figure 1 (one is hidden due to overlap).

The water depth (the dependent variable of the models) and the radiances at eight spectral bands (the independent variables of the models) were analysed for validity of normality. For this purpose, Q-Q plots were prepared assuming normal, log-normal and exponential distributions. These plots were used to determine the form (as is, logarithmic transformation, or exponential transformation) of the independent variables that would be used in regression model development. The distribution type of a variable was selected based on the slope information in the corresponding Q-Q plot. If the slope was close to 1, the corresponding distribution type was selected for the given variable. Following the determination of independent variable distribution forms, multicollinearity analysis was performed in an iterative procedure to identify highly correlated independent variables. At this stage, correlation matrix was used to remove an independent variable that had the highest correlation with another. Then, a new correlation matrix was generated for the remaining variables. This cycle was repeated until multicollinearity was eliminated between independent variables. Correlation coefficient (r) was used as the criterion for variable elimination. It was assumed that multicollinearity existed between variables if the absolute r value was greater than 0.6. The remaining variables proceeding multicollinearity analysis were considered in regression model development in prediction of the water depths or bathymetry of Lake Eymir.

Performances in bathymetry determinations using multispectral images are variable for Case 1 and Case 2 waters. In Case 1 waters (i.e., open ocean) chlorophyll is the main optically active constituent. These waters generally lack suspended particles. On the other hand, there is a complex relationship between reflectance and water quality parameters in Case 2 waters (i.e., coastal, estuary or inland waters such as lakes). This complexity is mainly due to the co-presence of chlorophyll, suspended particles and coloured dissolved organic matter in high concentrations (Kishino et al. 2005; Sudheer et al. 2006). Since depth determination in Case 2 waters or turbid lakes using multispectral images can be problematic, data elimination may be required to improve the prediction capability of the bathymetry models. Stevens (1984) showed that regression model prediction performance can be improved by eliminating outlier data based on standardized residuals (SR) of a regression model. In this approach, first a regression model is developed using the data set. Then, outlier observation points are determined and a new model is developed using the remaining observation points. In this study, a similar approach was used to eliminate the outlier observations. First, an initial regression model was developed. Then, the standard deviation of SR (SD_S) was calculated. At this stage, different multipliers (n) of SD_S were evaluated (n = 1.5, 1.4, ..., 0.5). Then, observation data with an SR greater than n × SD_S were eliminated. For each case (n × SD_S), a linear regression model was created. Then, these models were assessed based on basic statistics (minimum, maximum, mean and standard deviation) for the dependent variable (depth), and F-test.
for predictions. The model with a small F-value and basic statistics similar to the original observation data (48 observation points) was chosen as the best model. For the observation data for Lake Eymir, \( n = 0.7 \) resulted in the best filter in establishment of the data set for model development. However, this filter \((0.7 \times SD\_SR)\) resulted in elimination of 23 additional sampling points. As a result, bathymetry model developments were realized using the 25 sampling points depicted in Figure 1, which corresponded to a sampling density of 20 samples per square kilometre of the lake. Thirty-two per cent of these points (eight sampling points) were used in the model development stage. The remaining 68\% (17) were employed for model validation. Allocation of the locations of the sampling points for model development and model validation was performed arbitrarily while care was taken to have as even a spatial distribution as possible.

Following data screening, linear and non-linear regression models were developed to predict the bathymetry of the lake. The general forms of the linear and non-linear regression models are given in Equations (1) and (2), respectively:

\[
d_i = a + \sum_{j=1}^{J} k_j x_{ij}
\]

\[
d_i = \sum_{j=1}^{J} k_j x_{ij}^{m_j}
\]

where, \( d_i \) is the water depth at location \( i \), \( a \) is the intercept, \( k_j \) is the regression coefficient for band \( j \), \( x_{ij} \) is the radiance at location \( i \) at band \( j \), and \( m_j \) is the exponent for band \( j \). The \( a \) represents the offset for the depth of 0 m \((Loomis 2009)\) for the linear regression model. This parameter is used to handle the average error that would be produced by over- and under-predictions at different depths as a result of the impact of heterogeneous bottom cover (macrophytes, sand, gravel, etc.) and variable water quality (suspended solids, chlorophyll, etc.) on reflectance values \((Loomis 2009)\) for the linear model. In the above equations, \( a \), \( k_j \) and \( m_j \) values are set by XLStat software by minimizing the root mean square error (RMSE) and maximizing the Pearson coefficient of determination \((R^2)\) values between observed and predicted water depths at given locations.

**RESULTS AND DISCUSSION**

The Q–Q plots indicated that the radiance data in most of the spectral bands had normal distributions, except in Band 5 and Band 2. In these bands, log-normal distributions prevailed. Based on this information, the log transformations (base 10) of the data in Bands 5 and 2 were used in multicollinearity and correlation analysis. The correlation matrix indicated that Band 1 (coastal blue) was highly correlated with Bands 2, 3 and 4 \((r > 0.75)\). Moreover, at 95\% confidence level, \( r \) values for the relationships between Band 1 and Band 6, and Band 1 and Band 7 were higher than 0.6, which was the lower limit for multicollinearity elimination. Multicollinearity analysis indicated that only the data in Band 1, Band 8 and the logarithmic transform of the data in Band 5 were independent from each other and could be used as explanatory variables in regression model development. Puetz et al. (2009) and Maheswari (2010) showed the usefulness of inclusion of Bands 1 and 8 in depth determinations in coastal waters as well.

Band 1 senses the radiation in the 400–450 nm wavelength interval. This band supports bathymetric studies by sensing the deeper parts of a water body compared to other sensors (Puetz et al. 2009). Band 5, on the other hand, acquires radiance data in the range of 630–690 nm. The light in this region of the electromagnetic spectrum is mainly absorbed by chlorophyll-a \((Thiemann & Kaufmann 2000)\). As mentioned before, analysis of the data in this band revealed a log-normal distribution. This was in line with the distribution of measured chlorophyll-a concentrations in Lake Eymir. The radiance in Band 8 \( (x_{i8}) \) was another explanatory variable that was used in bathymetry model development for Lake Eymir. In various studies, the relationship between suspended particles and radiance in near-infrared band has been shown \((Doxaran et al. 2002)\). In this study, the impact of suspended particles in depth determination was considered through inclusion of Band 8. An initial analysis of the distribution of suspended particle concentrations in Lake Eymir indicated a
normal distribution similar to that for the distribution of radiance values in Band 8. It must also be noted that it is possible to observe frequent algal blooms over the lake surface in patches. Moreover, macrophytes may cover the bottom, especially at shallower depths closer to shore. Therefore, reflectance in Band 8 can be impacted by these as well.

Preceding the determination of independent explanatory variables (that have no multicollinearity), data screening was performed. As mentioned earlier, the sampling points over re-suspension areas were removed from the data set in order to avoid the interference these areas would produce in depth predictions. Then, the remaining 48 sampling points were taken into consideration. The minimum, average and maximum water depths at these points were 2.50, 4.57 and 5.75 m, respectively. These values were 2.50, 4.56 and 5.75 m, respectively, for the full observation data set (59 observation points). Further data elimination was conducted based on SD\_SR. This approach was used to improve the prediction capability of depth models. Application of remote sensing technology to Case 2 waters to make water quality predictions may be problematic compared to Case 1 waters due to the presence of water constituents that may significantly impact radiance values (Swardika 2007). It is very probable that local algal blooms, bottom sediments, suspended particles, and even waves can impact radiance values. Another difficulty is the heterogeneous distribution of these interferences which may lead to extreme values. By regarding extreme values as outliers, the impact of such interferences in model prediction performance may be improved at least at other locations in the lake that are less prone to such effects. As seen in Figure 1, removed sampling points form clusters in certain locations. It is possible that these locations were subject to the interferences mentioned. The minimum, average and maximum depths for 25 observation points used in the model development were 2.80, 4.70 and 5.70 m, respectively. Therefore, it can be said that deeper locations were considered as ground truth data for modelling purposes. It may be the case that deeper locations impacted less from bottom sediment re-suspension or bottom reflection.

The linear and non-linear regression models generated to determine the depths at different locations using remotely sensed data are given below in Equations (3) and (4), respectively:

\[
d_i = -2.433 + 193.000x_{i1} - 1.313 \log x_{i5} - 108.886x_{i8} \tag{3}
\]

\[
d_i = 1140.027x_{i1}^{1.628} - 0.128 \log x_{i5}^{5.000} - 419.672x_{i8}^{1.378} \tag{4}
\]

\[R^2, \text{ adjusted } R^2, \text{ RMSE and } F\text{-value with respect to the calibration data set were 0.87, 0.78, 0.370 and } 6.78 \times 10^{-4}, \text{ respectively, for the linear regression model at 95% confidence level. For the same data set, } R^2, \text{ adjusted } R^2, \text{ RMSE and } F\text{-value were 0.90, 0.83, 0.379 and } 3.04 \times 10^{-4}, \text{ respectively, when the non-linear regression model was used. Performances of these models were also tested against the validation data. } R^2, \text{ adjusted } R^2, \text{ RMSE and } F\text{-value were 0.805, 0.760, 0.488 and } 1.07 \times 10^{-6}, \text{ respectively, when the linear regression model was applied. The corresponding values for the non-linear model were 0.855, 0.822, 0.365 and } 1.11 \times 10^{-7}, \text{ respectively, at 95% confidence level. In both models, the radiance values in Band 1 had the highest coefficient (} k_i \text{ in Equations (1) and (2)) compared to other bands, keeping in mind that the radiance in Band 5 (} x_{i5} \text{) was in logarithmic scale. This situation emphasized the importance of Band 1 in bathymetry determination. A similar observation was valid in the correlation matrix as well. Compared to other bands, depth had the highest } r \text{ (0.351) for Band 1 radiance at 95% confidence level when 48 sampling points were considered. The coefficients for Bands 5 and 8 were negative which were indicative of the interference due to absorption based on the presence of suspended solids and algal species. It must also be noted that another model based on the ratio method proposed by Stumpf et al. (2005) was tested. The ratio of } \ln(x_{i5})/\ln(x_{i1}) \text{ was used. This ratio had the highest correlation with depth } (R^2 = 0.51) \text{ compared to other combinations. The model obtained for this ratio, } d_i = 15.652*(\ln x_{i5}/\ln x_{i1}) - 11.886, \text{ resulted in no better performance. } R^2 \text{ and RMSE were 0.46 and 0.529 at 95% confidence level.}

Measured versus predicted depths are depicted in Figure 3. As can be seen, both models were successful in predicting low and high depths in Lake Eymir. However, the statistical analysis given before stated that the non-linear model was slightly better in depth predictions. For the
validation data set, the average error was calculated as 0.30 and 0.2 m for the linear and non-linear regression models, respectively. The average depth for the validation data set (observations) was 4.73 m. The predicted average depths were 4.65 and 4.71 m for the linear and non-linear regression models, respectively. These correspond to 2 and 0.5% error in the predicted average depths, respectively. Therefore, models developed using screened ground truth data were successful in predicting the average depth. When the models were applied to predict the depths at 48 sampling points, the average error in depth predictions was 0.61 m in the linear regression model and 0.55 m in the non-linear regression model. The errors in the calculated average depths were 13 and 12%, respectively, for the linear and non-linear models.

The bathymetric maps of Eymir Lake that are generated using Equations (3) and (4) are depicted in Figures 4 and 5, respectively. Both models simulated the shallow depths at shores with success. The increasing depths from the shoreline can be clearly seen for both models. Tureli & Norman (1992) studied the bathymetry of the lake using sonar technology. According to that study, the lake bottom had a bowl-type structure with steep slopes at shores. As a result, sharp increases were observed in depths progressing away from the shore to the inner regions of the lake. The mid-region of the lake was the deepest location with an average depth of 5.5 m in 1985. They also stated that the lake became relatively shallow at the southern and eastern parts, which correspond to the inlet and outlet of the lake, respectively. The findings of Tureli & Norman (1992) are consistent with the results of this study. As Lake Eymir has a valley-type structure, a sharp increase is expected in depth in short distances away from the shore. This is captured by the depth models (Figures 4 and 5). Moreover, the southern and eastern parts of the lake are shallower than the other parts. The deeper regions of the lake are shown by darker shades in Figures 4 and 5. In general, the
distributions of relatively lower and higher depths were in line with the observations. However, the depths at re-suspension areas were in error. This could be seen especially at the southern part of the lake closer to the inlet. At these locations mixed values were observed. Overall, although both models predicted the depths well, the non-linear model was better in predicting the shallower depths at shores. However, the non-linear model was more sensitive to the impact of re-suspension areas.

CONCLUSIONS

The results of this study showed that WorldView-2 image can be used to predict the depths in a eutrophic lake. Bands 1, 5 and 8 of the WorldView-2 satellite were adequate to determine the depth distribution. Among these bands, Band 1 (coastal blue band) made the highest contribution in determination of the depths in the eutrophic lake.

The presence of turbidity due to re-suspension areas caused interference in predicting the depths. However, eliminating these areas in the depth model development helped to make good depth estimates at locations where the impact of turbidity was less. More study is required to deal with this issue and improve the prediction capability of the models at these locations as well. For the existing situation, regression models were successful in defining the shallow depths at shore and close to the inlet and outlet of the lake. Moreover, deeper locations were successfully identified.

Bathymetry determination using WorldView-2 can aid in water quality studies. Use of remotely sensed data may provide an alternative in determination of the distribution of depths and examination of the water quality in lakes with respect to these depths. Scale advantage supplied by remote sensing over traditional bathymetry generation methods may make it preferable for large lakes. However, more research is needed to investigate the effects of spatially and temporarily heterogeneous bottom characteristics (i.e., variable coverage by macrophytes, different bottom materials) on reflectance values in determination of depths in a eutrophic lake.

ACKNOWLEDGEMENTS

The authors are grateful to the Scientific and Technical Research Council of Turkey (TUBITAK) for providing financial support for this study (Project Number: CAYDAG-106Y201). The authors acknowledge Res. Asst. Elif Kucuk for her support during field work.
REFERENCES


First received 1 August 2012; accepted in revised form 9 May 2013. Available online 6 June 2013