Deriving Lake Urmia water depth using Landsat 8 imagery and field measurements

Yasmin Ghadyani [1]

Abstract

Accurate estimation of the volume of water in Lake Urmia plays a crucial role in evaluating the effectiveness of the implemented lake restoration plans. This study aims to find a detailed depth map inside the lake using in situ measurements of the lake bed elevation conducted in 2018.

Several models have been used to derive the relationship between the Landsat 8 imagery reflectance values in different bands and the measured water depth. The best model, a random forest regressor, can predict the test data with a coefficient of determination and a mean squared error of 0.8612 and 0.02m, respectively. In the next step, the developed method is used to predict water depth of the lake.

1. Introduction

Many saline lakes worldwide have witnessed extreme desiccation over the past decades, e.g., the Aral Sea (located between Kazakhstan and Uzbekistan) (Micklin, 1988) and (Micklin, 2007), Lake Urmia (LU) (Iran) AghaKouchak et al., 2015), Great salt Lake (Utah, USA) (Wurtsbaugh et al., 2016) and Walker Lake (Nevada, USA) (Beutel et al., 2001). This endangering condition has severely affected water quality, human and biota health, and regional economics (e.g., AghaKouchak et al., 2015)).

Since anthropogenic activities (such as extensive agriculture development) have imposed a much more significant contribution to drying than climate variability (e.g., Wurtsbaugh et al. 2017), effective management practices for the urgent and sustainable restoration of these environments are indispensable. Monitoring the fluctuation of water storage in a reviving lake is one reliable way to assess the success of any restoration plans. This can be accomplished by estimating the water balance components, i.e., direct precipitation on the lake, surface and groundwater inflows, evaporation rate, and other possible forms of water loss. However, this is challenging due to the complexity and uncertainties involved in estimating these variables.

Alternatively, water storage can be estimated indirectly from the geometry of a water body, requiring the knowledge of the bed topography or bathymetry. The bathymetry of aquatic environments carries essential information about the estuarine circulation (Simionato et al., 2004), delta morphology (Brucker et al., 2007), reef ecology (Wedding et al. 2008), tidal currents (Prandle 2003), wave propagation (Lynett and Liu 2002), bottom currents in the coastal zones, lake or reservoir sedimentation, and erosion-sedimentation rates (Dost and Mannaerts 2008). The bathymetry of a water body can be highly dynamic in both space and time due to the fluctuating sediment loads from the receiving rivers or tributaries. Recursive bathymetric monitoring of these environments is thus necessary to decipher how the physical processes controlling the bed dynamics are influenced by climate change and human interventions (Ceyhun and Yalçin 2010).

Bathymetric maps are traditionally derived directly by sonar measurements using single or multi-beam echo sounders (Sánchez-Carnero et al., 2012). Despite the high accuracy and unparalleled lake bottom coverage, this technique is not efficient in shallow waters. It is expensive, time-consuming, and labor-intensive, and there is a risk of echo sounders stranding (Smith and Sandwell 2004). Instead, Airborne LiDAR bathymetry (ALB) can characterize the depth of a water column in relatively shallow, clear water with high accuracy from an airborne platform using a scanning and pulsed light beam (Saylam et al., 2018). Nevertheless, the extremely high cost of ALB limits its application (Sánchez-Carnero et al., 2012). Alternatively, Satellite-Derived Bathymetry (SDB) by optical remote sensing is a low-cost, wide-coverage, and time-effective approach that relies on passive multispectral scanner data and uses optical characteristics of water column to frequently estimate water depth over a large spatial extent (Ceyhun and Yalçin 2010). SDB-based approaches can be categorized into analytical, semi-analytical, and empirical methods (Evagorou et al., 2019). Implementing data-driven empirical models such

[*]Equal contribution [1]Faculty of Geosciences and Environment, University of Lausanne, Lausanne, Switzerland. Correspondence to: Yasmin Ghadyani <yasmin.ghadyani@unil.ch>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

as the log-linear relationship between surface reflectance and water depth (Lyzenga, 1978) and the ratio transform relationship of Stumpf (Stumpf, Holderied, and Sinclair, 2003) are less challenging than the analytical and semi-analytical methods. This is because the complex influence of atmospheric properties, lake bed composition, and vegetation cover are lumped into a few parameters that can be calibrated using the available in-situ observations (Jagalingam, Akshaya, and Hegde 2015). Recently, advanced Machine Learning (ML) techniques coupled with satellite imagery have been successfully implemented in deriving lake bathymetry. These include, but are not limited to, Artificial Neural Networks (ANN) (Ceyhun Yalçin, 2010), (Gholamalifard et al., 2013), (Mohamed et al., 2017) and (Liu et al., 2018), Ensemble of Regression Trees using Bootstrap Aggregation (i.e., bagging) (Mohamed et al., 2017), Ensemble of Regression trees using least-squares boosting (LSB) (Mohamed et al., 2016) and (Mohamed et al., 2017), Random Forest (RF) (Manessa et al., 2016) and (Sagawa et al., 2019), and Support Vector Machine (SVM) algorithm (Misra et al., 2018) and (Wang et al., 2019).

This study has predicted the water depth map of LU with the help of the field measurements of the lake, and with the help of satellite imagery, using a random forest regressor model.

The rest of the paper is organized as follows. Section 2 describes an overview of the research, introducing the study area, the dataset, and different methods for deriving the water depth. In section 3, four different methodologies are studies to fidn the method that is able to best model the water depth, based on the bands 2 to 7 of the Landsat 8 imagery of the same area. Finally, using the best model, the water depth map for the data collection period is acquired. Finally, the fifth chapter sums up the conclusion of the study.

2. Case Study

LU is a hypersaline lake in the northwest of Iran (located between 44° 56’ E to 46° 04’ E longitudes and 36° 56’ N to 38° 22’ N latitudes) that was known as the second hypersaline lake worldwide before losing a vast majority of its water volume between 1995 to 2017 (AghaKouchak et al., 2015). Intensive agricultural development, the imbalance between water supply and demand, disrespecting the environmental demand of the lake demand by the reservoirs on the rivers feeding the lake, illegal withdrawal from surface and groundwater resources, and low irrigation and agricultural productivity in the Lake Urmia Basin resulted in the decline of LU’s water level by nearly eight meters (i.e., losing 96 percent of water storage from 34.22 BCM to 1.24 BCM) in this period (see Danesh-yazdi and Ataie-Ashtiani 2019 and references therein).

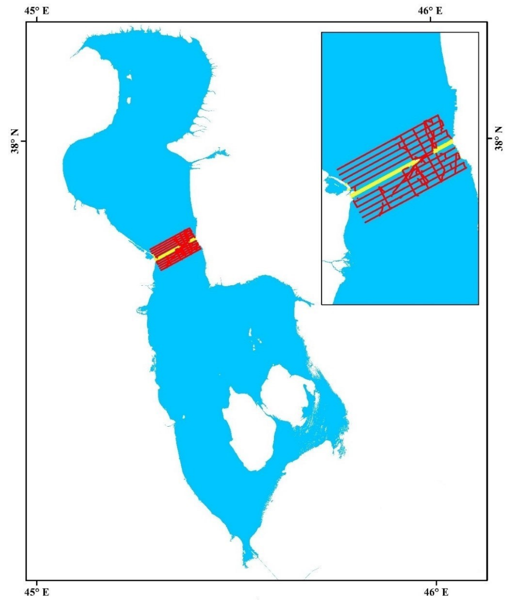

Figure 1. Path of the measurement profiles captures in Marc to April, 2018

This environmental catastrophe led to a wide range of environmental and socioeconomic problems such as desertification, an increase in soil salinity in agricultural areas near the lake, disturbance in the ecological condition of species living in the lake, initiation of salt storms from the dried portion of the lake, increase in health-related problems, and increase of unemployment rate due to recession in agricultural activities (Danesh-yazdi and Ataie-Ashtiani 2019)

3. Methodology

This section describes the data used and machine learning methodologies used to model it.

3.1. Data

3.1.1. Satellite imagery

For this study, Landsat 8 imagery with a spatial resolution of 30m, captured on 25 March 2018, has been used. , and the level 2 images containing surface reflectance (SR) values were converted from the top of atmosphere reflectance values using the LaSRC algorithm developed by Vermote (U.S. Geological Survey 2019).

3.1.2. Hydrography data

The water depth and bed elevation data acquisition campaigns were carried out from 6 March to 20 April 2018 across the LU by the Urmia Lake Restoration Program (ULRP) and the Geological Survey and Mineral Explorations of Iran (GSI). Figure 1 shows the path of these

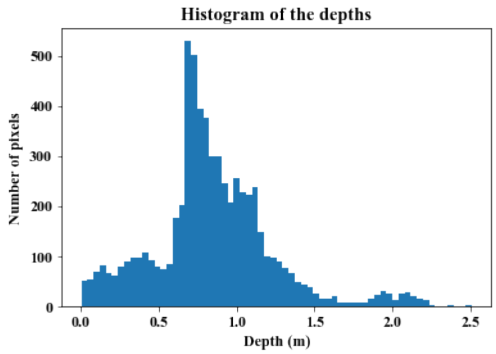

Figure 2. Histogram of all input data based on the depth (m).

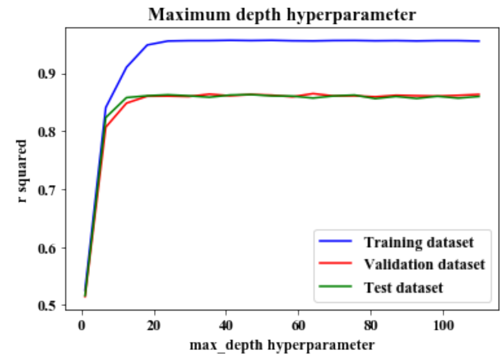

Figure 3. The r-squared value of the training, validation and testing algorithm for a sample random forest regressors with the following hyperparameters and a variable maxdepth between 1 and 110.

bathymetric surveys, extending in both the northern and southern parts of the lake. The measurements are done as 12 paths with a length of 10-km, which are centered around the Kalantari bridge. 500 m distances separate the paths. The water was acquired using daily LU water level data recorded at the Golmankhaneh station (37° 36 00.9 N and 45° 15 25.9 E) to compute water depth from the measured lake bed elevation data. In total, there are 6426 data, each corresponding to a measurement point in Figure 1, of which 60 percent is used for the training data and 20 percent used for validation and test dataset each.

3.2. Modeling

In this study, four methods have been used to model the depth using band reflectance values: Linear and Polynomial regressions, Regression SVM, and Random Forest Regressor.

4. Results

In this section, at first, different regression methodologies are explored to find the model that is able to best describe the three splits of data: training, validation and testing. Finally, the best model is used to map the water depth across the whole lake to find the depth in different locations.

4.1. Modeling

In order to describe the data using regression methods, the data has been split into three sets of training, validation and testing. The modeling starts with the simplest model, linear regression. The results of all the models have been provided in Table 1.

As the results show, this simple model is able to provide a model with a low mean squared error. However, the low r squared value, even for the training data shows the model does not yield the prefect result.

As a result, it has been decided to move to the more complex models which as the results show, are able to better explain the data. However, a pitfall of moving into the complex models is that their higher accuracies might come at the cost of overfitting, i.e. low performance on the validation and test sets.

As the result shows, by an increase in the complexity of the model, the difference in the performance between the training and test sets increases. However the test data is also yielding better results, moving from an r squared of 0.53 in the linear regression to an r squared of 0.86 for the random forest regressor method.

It should also be mentioned that the hyperparameters of all the studied methods have been relatively tuned to increse their performance. As an exmaple, for the case of linear SVM, different values for the parameters of kernel, degree, C and epsilon have been tested, reaching to the fine tuned values of ”rbf”, 2, 1000, and 0.1, respectively.

Samely, for the case of RF-R, testing different hyperparameters, as shown in Table 2, gives the best results for the following parameters: nestimators = 1400, minsamplessplit = 5, minsamplesleaf = 1, maxfeatures = ’sqrt’, maxdepth = 30, bootstrap = True.

Figure 3 shows the r squared values for different values of maxdepth for training, validation, and test datasets. As the figure shows, which is comptible with reults of Table 1, the training value is able to yield better results compared with the test and validation data, for which the results are the same. However, it interestingly shows the effect of tuining hyperparameters (maxdepth as a sample), As it shows, after an approximate value equal to 30, the increase of this parameter will not chnage the results. However, for values less than this threshold, this parameter plys an important role in determining the results.

Table 1. Regression metrics, r squared value and accuracies for the four tested algorithms for training and testing algorithms: Linear regression (LR), Plynomial linear regression (PLR), regression SVM (R-SVM), and random forest regressor (RF-R).

METHOD	R^2_{train}	$ACCU_{train}$	R^2_{test}	$ACCU_{test}$
LR	0.549	0.070	0.531	0.071
PNR	0.685	0.049	0.662	0.052
R-SVM	0.812	0.028	0.780	0.034
RF-R	0.957	0.007	0.861	0.021

Table 2. Hyperparameters used for tuning the random forest regressor models.

HYPERPARAMETER	VALUES
$\mathrm{N}_e stimators$	200, 400, 600, ..., 1600, 1800, 2000
$\mathrm{MAX}_f eatures$	AUTO, SQRT
$\mathrm{MAX}_d epth$	10, 20, 30, ..., 90, 100, 110
$\mathrm{MIN}_s amples_s plit$	2, 5, 10
$\mathrm{MIN}_s amples_l eaf$	1, 2, 4
BOOTSTRAP	TRUE, FALSE

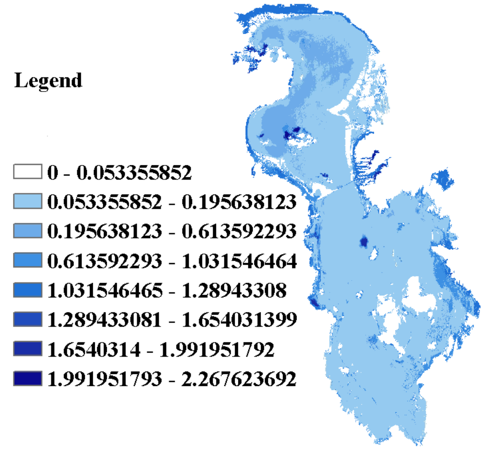

Figure 4. The calculated depth of Lake Urmia for the date of study, 25 March 2018, based on the final random forest reressor model and Landsat 8 imagery.

4.2. Prediction

In this section, the selected model , the random forest regressor with tuned hyperparameters is used to predict the water depth in the lake. Figure 4 shows the water depth map. It should be mentioned that the water is delineated using the unsupervised k-means algorithm in MATLAB.

Based on the results, the depth values range between the values of zero and 2.3 meters. The maximum value happens mostly in the margins of the lake and also in the center of the northern and southern parts.

However, based on the field observations, I assume that getting high values in the margins of the lake might not be very true. I presume this might be a result of a problem in water delineation. As the delineated water masks can significantly change the results, I presume enhancing the water delineation algorithm can be interesting and might give more reasonable result.

5. Conclusion

In this study, several regression methods: linear regression, polynomial linear regression, regression SVM, and random forest regressor, have been used to model water depth in Lake Urmia using Landsat 8 satellite imagery. If possible, the hyperparameters are tuned in each case to acquire the best result.

The results show that the RMSE is not very high, even for the simplest model. In contrast, the R^2 is low not only for the test but also for the training dataset. As the model becomes more and more complex, the R^2 increases and therefore, the most complex model can give the best results.

The model is able to predict reasonable (at least not unreasonable!) results for the lake depth, which is an advantage. However, the performance of the model is highly affected by the water mask in this case, which could be enhanced.

The advantage of this model is that compared with other more complex models, i.e. ANN is much simpler and still, is able to describe the data excellent performance metrics.

Software and Data

In this project, the Landsat 8 imagery has been downloaded from the USGS website, accessible via this link. The confidential field measurement data have been provided by the Urmia Lake Restoration Program (ULRP). The field measurements and the satellite imagery have been processed in MATLAB to extract the relevant information. The final dataset, including the inputs and outputs used in this study, has been made accessible in the following link. The data is then processed by python, and the notebook is accessible on GitHub through this link.

References

AghaKouchak, Amir et al. 2015. "Aral Sea Syndrome Desiccates Lake Urmia: Call for Action." Journal of Great Lakes Research 41(1): 307–11.

Beutel, Marc W, Alex J Horne, James C Roth, and Nicola J Barratt. 2001. "Limnological Effects of Anthropogenic Desiccation of a Large, Saline Lake, Walker Lake, Nevada." In Saline Lakes, Springer, 91–105.

Brucker, Steve et al. 2007. "Monitoring Flood-Related Change in Bathymetry and Sediment Distribution over the Squamish Delta, Howe Sound, British Columbia." In US Hydro. Conf,.

Burges, Christopher J C. 1998. "A Tutorial on Support Vector Machines for Pattern Recognition." Data Mining and Knowledge Discovery 2: 121–67.

Ceyhun, Özçelik, and Arisoy Yalçin. 2010. "Remote Sensing of Water Depths in Shallow Waters via Artificial Neural Networks." Estuarine, Coastal and Shelf Science 89(1): 89–96.

Danesh-yazdi, Mohammad, and Behzad Ataie-Ashtiani. 2019. "Lake Urmia Crisis and Restoration Plan: Planning without Appropriate Data and Model Is Gambling." Journal of Hydrology 576(June): 639–51.

Dost, R J J, and C M M Mannaerts. 2008. "Generation of Lake Bathymetry Using Sonar, Satellite Imagery and GIS." In Proceedings of the 2008 ESRI International User Conference: GIS, Geography in Action. San Diego, Florida,.

Evagorou, Evagoras et al. 2019. "Bathymetric Maps from Multi-Temporal Analysis of Sentinel-2 Data: The Case Study of Limassol, Cyprus." Advances in Geosciences 45(1988): 397–407.

Geyman, Emily C., and Adam C. Maloof. 2019. "A Simple Method for Extracting Water Depth From Multispectral Satellite Imagery in Regions of Variable Bottom Type." Earth and Space Science 6(3): 527–37.

Gholamalifard, Mehdi et al. 2013. "Remotely Sensed Empirical Modeling of Bathymetry in the Southeastern Caspian Sea." Remote Sensing 5(6): 2746–62.

Jagalingam, P., B. J. Akshaya, and Arkal Vittal Hegde. 2015. "Bathymetry Mapping Using Landsat 8 Satellite Imagery." Procedia Engineering 116(1): 560–66.

Karimi, Neamat, and Mohammad Hossein Bagheri. 2016. "Deriving and Evaluating Bathymetry Maps and Stage Curves for Shallow Lakes Using Remote Sensing Data." Water Resources Management.

Liu, Shan et al. 2018. "Deriving Bathymetry from Optical Images with a Localized Neural Network Algorithm." IEEE Transactions on Geoscience and Remote Sensing 56(9): 5334–42.

Lynett, Patrick J, and Philip L.-F. Liu. 2002. "A Two-Dimensional, Depth-Integrated Model for Internal Wave Propagation over Variable Bathymetry." Wave Motion 36(3): 221–40.

Lyzenga, David R. 1978. "Passive Remote Sensing Techniques for Mapping Water Depth and Bottom Features." Applied Optics 17(3): 379.

MacQueen, J. 1967. "Some Methods for Classification and Analysis of Multivariate Observations." In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, Berkeley, Calif.: University of California Press, 281–97.

Manessa, Masita Dwi Mandini et al. 2016. "Satellite-Derived Bathymetry Using Random Forest Algorithm and Worldview-2 Imagery." Geoplanning: Journal of Geomatics and Planning 3(2): 117.

Micklin, Philip. 1988. "Desiccation of the Aral Sea: A Water Management Disaster in the Soviet Union." Science 241(4870): 1170–76. ———. 2007. "The Aral Sea Disaster." Annual Review of Earth and Planetary Sciences 35(1): 47–72.

Misra, Ankita, Zoran Vojinovic, Balaji Ramakrishnan, and Arjen Luijendijk. 2018. "Shallow Water Bathymetry Mapping Using Support Vector Machine (SVM) Technique and Multispectral Imagery Shallow Water Bathymetry Mapping Using Support Vector Machine (SVM) Technique and Multispectral Imagery." International Journal of Remote Sensing 00(00): 1–20.

Mohamed, Hassan et al. 2017. "Assessment of Proposed Approaches for Bathymetry Calculations Using Multispectral Satellite Images in Shallow Coastal/Lake Areas: A Comparison of Five Models." Arabian Journal of Geosciences 10(2): 1–17.

Mohamed, Hassan, Abdelazim Negm, Mohamed Zahran, and Oliver C. Saavedra. 2016. "Bathymetry Determination from High Resolution Satellite Imagery Using Ensemble Learning Algorithms in Shallow Lakes: Case Study El-Burullus Lake." International Journal of Environmental Science and Development 7(4): 295–301.

Prandle, D. 2003. "Relationships between Tidal Dynamics and Bathymetry in Strongly Convergent Estuaries." Journal of Physical Oceanography 33(12): 2738–50.

Sagawa, Tatsuyuki, Yuta Yamashita, Toshio Okumura, and Tsutomu Yamanokuchi. 2019. "Satellite Derived Bathymetry Using Machine Learning and Multi-Temporal Satellite Images." Remote Sensing 11(10): 1155.

Sánchez-Carnero, N. et al. 2012. "Fast and Low-Cost Method for VBES Bathymetry Generation in Coastal Areas." Estuarine, Coastal and Shelf Science 114: 175–82.

Saylam, Kutalmis et al. 2018. "Airborne Lidar Bathymetry: Assessing Quality Assurance and Quality Control Methods with Leica Chiroptera Examples." International Journal of Remote Sensing 39(8): 2518–42.

Sima, S, and M Tajrishy. 2013. “Using Satellite Data to Extract Volume – Area – Elevation Relationships for Urmia Lake , Iran.” Journal of Great Lakes Research 39(1): 90–99.

Simionato, C G, W Dragani, V Meccia, and M Nuñez. 2004. “A Numerical Study of the Barotropic Circulation of the Rıo de La Plata Estuary: Sensitivity to Bathymetry, the Earth’s Rotation and Low Frequency Wind Variability.” Estuarine, Coastal and Shelf Science 61(2): 261–73.

Smith, Walter H.F., and David T. Sandwell. 2004. “Conventional Bathymetry, Bathymetry from Space, and Geodetic Altimetry.” Oceanography 17(SPL.ISS. 1): 8–23.

Stumpf, Richard P., Kristine Holderied, and Mark Sinclair. 2003. “Determination of Water Depth with High-Resolution Satellite Imagery over Variable Bottom Types.” Limnology and Oceanography 48(1 II): 547–56.

U.S. Geological Survey. 2019. “Landsat 8 Surface Reflectance Code (LASRC) Poduct Guide. (No. LSDS-1368 Version 2.0).” (May): 40. https://www.usgs.gov/media/files/landsat-8-surface-reflectance-code-lasrc-product-guide.

Wang, Lei, Hongxing Liu, Haibin Su, and Jun Wang. 2019. “Bathymetry Retrieval from Optical Images with Spatially Distributed Support Vector Machines.” GIScience and Remote Sensing 56(3): 323–37. https://doi.org/10.1080/15481603.2018.1538620.

Wedding, Lisa M et al. 2008. “Using Bathymetric Lidar to Define Nearshore Benthic Habitat Complexity: Implications for Management of Reef Fish Assemblages in Hawaii.” Remote Sensing of Environment 112(11): 4159–65.

Wurtsbaugh, Wayne et al. 2016. “Impacts of Water Development on Great Salt Lake and the Wasatch Front.” Watershed Sciences Faculty Publications: 1–9. ———. 2017. “Decline of the World’s Saline Lakes.” Nature Geoscience 10(11): 816–21.