ESTIMATING WATER LEVEL IN THE URMIA LAKE USING SATELLITE DATA: A MACHINE LEARNING APPROACH

M. Boueshagh 1, M. Hasanlou 1,*

1 School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran (booesghah.mahboob, hasanlou)@ut.ac.ir

KEYWORDS: Water Level, Urmia Lake, MODIS Snow Cover, Evaporation, Precipitation, Support Vector Regression (SVR), Water Budget Procedure

ABSTRACT:
Lakes play a pivotal role in the development of cities and have major impacts on the ecosystem balancing of the area. Remote sensing techniques and advanced modeling methods make it possible to monitor natural phenomena, such as lakes’ water level. The ecosystem of Urmia Lake is one of the most momentous ecosystems in Iran, which is almost close-ended and has become a global environmental issue in recent years. One of the parameters affecting this lake water level is snowfall, which has a key role in the fluctuations of its water level and water resources management. Hence, the purpose of this paper is the Urmia Lake water level estimation during 2000-2006 using observed water level, snow cover, direct precipitation, and evaporation. For this purpose, Support Vector Regression (SVR), which is the most outstanding kernel method (with various kernel types), has been used. Furthermore, four scenarios are considered with different variables as inputs, and the output of all scenarios is the water level of the lake. The results of training and testing data indicate the substantial impact of snow on retrieving the water level of the Urmia Lake at the desired period, and due to the complexity of the data relationships, the Gaussian kernel generally had better results. On the other hand, Quadratic and Cubic kernels did not work well. The fourth scenario, with RBF kernel has the best results [Training: $R^2 = 97\%$ and RMSE = 0.09 m, Testing: $R^2 = 96.97\%$ and RMSE = 0.08 m].

1. INTRODUCTION
Nowadays, some lakes throughout the world are drying or completely dried due to irrigation or urban use diverting inflow. Climate changes are also a factor in drying some lakes (CASEY, 2016). Urmia Lake is one of the largest permanent lakes in the Middle East with a catchment area of 51,876 km², comprising 3% of the country area and 7% of total surface water in Iran, which has started serious uninterrupted desiccation since 1995 (Figure 1). There are many reasons for this dramatic water losses: (1) the thirteen rivers entering the lake have been dammed, (2) increased groundwater pumping has reduced flows into the lake, (3) water diversions, (4) climate changes, and (5) drought (Eimanifar et al., 2007), (Marks, 2019). Naturally, the most important factors affecting the water level of this lake are climatic and hydrological factors. For example, precipitation and evaporation can be one of the most important climatic factors that are considered in this study. As a result, prediction of water level and investigating its future trends are crucial to prevent the complete drying of the lake. During 2002-2005, even though the water entering the lake (including permanent rivers and precipitation on the lake) is less than the threshold of Urmia Lake’s evaporation, there has been a temporary increase in the water level of Urmia Lake that can be due to the temporary runoff from snow melting (Boueshagh et al., 2019). If this enhancement is due to the temporary runoff from snow melting, and besides, the amount of water entering the lake is greater than the threshold of evaporation, or at any time, the water entering the lake is not less than the threshold of evaporation, then the lake will have no shrinkage. Thus, with any rainy year and the snowmelt runoff, the Urmia Lake will naturally be recovered. Therefore, snow in the catchment area of Urmia Lake is one of the significant sources of the lake water, and Snow Cover (SC) plays a significant role in the water balancing. Because of these reasons, SC monitoring is an important tool for analyzing this lake’s water level, particularly because satellite data provide timely and efficient snow cover information for large areas. Estimation of SC and water stored in the snowpack has great importance in climate changes researches and water resources management (Nikefatari et al., 2016), (Johnson et al., 1984), (Yasunari et al., 1990), (Cohen, 1994). According to the studies, snow is a controlling parameter over the regional weather and climate patterns (Cohen et al., 1999). Evaluation of long-lasting SC can affect climate changes monitoring as well as the simulating models (Frei et al., 1999). On a regional scale, SC is essential for local water availability, river run-off, and groundwater recharge, especially in middle and high latitudes (AKYÜREK et al., 2002). In countries such as Norway or Switzerland, where electricity is mainly generated through hydropower stations, SC is an important factor influencing energy supply (Vikhamar et al., 2003). In this context, exact knowledge about the snow-covered area is also essential for water resource management e.g., when using snowmelt runoff models (Butt et al., 2011).

* Corresponding author
Spatial-temporal attributes of SC can be obtained indirectly from remotely sensed data, or directly from ground measurements. In the remote sensing, one of the sensors that can be used to appraise SC is the Moderate Resolution Imaging Spectroradiometer (MODIS), and snow map is a SC estimation algorithm that was developed especially for MODIS (Hall et al., 1995). The MODIS daily, 8-day and monthly SC products provided by the National Snow and Ice Data Center (NSIDC) are based on the mentioned algorithm. It uses the reflectance of VIS and IR radiation to calculate the Normalized Difference Snow Index (NDSI):

\[
NDSI = \frac{b_4 - b_6}{b_4 + b_6}
\]

where \(b_4\) and \(b_6\) refer to MODIS bands 4 (0.54–0.56\(\mu m\)) and band 6 (1.62–1.65\(\mu m\)), respectively (Hall et al., 2002). Zhang et al. have used MODIS L1B data, MODIS Daily (MOD10A1) and 8-day (MOD10A2) Snow Products to monitor the SC of Liaoning Province in China over the winter months of November–April 2006–2008 (Zhang et al., 2010). There are many ways to examine the behavior of a parameter, as well as predict its future behaviors such as Support Vector Regression (SVR), Artificial Neural Network (ANN), and Least Squares (LS). In research, the potential of the support vector machine (SVM) in long-term prediction of lake water levels has been examined. They used the average value of monthly water levels from 1918 to 2001 to predict future water levels up to 12 months ahead and compared the results with a widely used neural network model called a Multi-Layer Perceptron (MLP) and with a conventional multiplicative Seasonal Autoregressive Model (SAR). Based on the results of this research (Khan et al., 2006), overall, the SVM showed good performance and is proved to be competitive with the MLP and SAR models. The SVM is relatively new and has demonstrated a good performance in classification (Osmun et al., 1997), (Belousov et al., 2002), regression (Smola et al., 2004), (Dibike et al., 2001), and time series forecasting and prediction (Mukherjee et al., 1997), (Müller et al., 1997), (Tay et al., 2001), (Kim, 2003), (Thissen et al., 2003). The SVM leads to a unique and global solution because of its formulation, which employs a structural risk minimization (SRM) principle as opposed to an empirical risk minimization (ERM) principle, employed by conventional neural networks (Dibike et al., 2001). The SRM places an upper bound on the expected risk, as opposed to an ERM, which minimizes the error on the training data only. In this paper, the SVR method, which is the most illustrious kernel method, with various kernel types has been used to estimate the Urmia Lake water level using measured water-level in the monthly average, monthly SC in the whole catchment area of the Urmia Lake, monthly direct precipitation and monthly direct evaporation of lake surface.

### 2. STUDY AREA AND DATASETS

The study area is the Urmia Lake, located in the northwest of Iran in a geographic position between 45°E - 46°E and 37°4N - 38°1’7N, is an oligotrophic lake of thalassohaline origin with a total surface area between 4750 and 6100 \(km^2\) and a maximum depth of 16 m at an altitude of 1250 m (Van Stappen et al., 2001) (Figure 2).

The total catchment area of the lake is about 51,876 \(km^2\), which is 3.15% of that of the entire country and includes 7% of the total surface water in Iran. There are thirteen main rivers in the lake basin, among them Zarrineh rood is the largest one with a total annual discharge value of about 2×109 \(m^3\) (Gaheri et al., 1999), (Touloie, 1998). The climate in the Urmia Lake basin is harsh and continental, affected mainly by the mountains surrounding the lake (Gaheri et al., 1999), (Alipour, 2006). Considerable seasonal fluctuations in air temperature occur in this semi-arid climate with an annual average precipitation of between 200 and 300 \(mm\). The air temperature usually ranges between 0°C and −20°C in winter and up to 40°C in summer. From this point of view, Urmia Lake is a critical asset for the region, because it acts to moderate these extremes (Kelts et al., 1986). Annual inflow into the lake is 6900×106 \(m^3\), of which 4900×106 \(m^3\) is from rivers, 500×106 \(m^3\) from floodwater (through rainfall) and 1500×106 \(m^3\) from precipitation (Touloie, 1998). Underground springs are also a source of water, but the volume is not known (Jalali, 1984). Measured water-level by tide gauges in the monthly average, monthly SC in the whole catchment area of the Urmia Lake from SC products with a resolution of 0.05 degree (approx. 5 km) related to the MODIS aboard the Terra satellite extracted from the National Aeronautics and Space Administration website (NASA) available online at: [https://earthdata.nasa.gov/](https://earthdata.nasa.gov/), monthly direct
precipitation of lake surface from the MSWEP (Multi-Source Weighted-Ensemble Precipitation) data set extracted from Princeton Climate Analytics (PCA) website available online at https://platform.princetonclimate.com/PCA_Platform/, and monthly direct evaporation from the ECMWF (European Centre for Medium-Range Weather Forecasts) model extracted from the European Centre for Medium-Range Weather Forecasts website available online at https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/, have been used for this study. This data is from 2000 to 2006, which there has been a temporary increase in the lake’s water level (Figure 3).

\[ Q_{\text{input}} = Q_{\text{output}} \pm \Delta V \]  

(2)

where \( Q_{\text{input}} \) and \( Q_{\text{output}} \) are the amounts of water entering the lake and the water coming out of the lake, respectively. The \( \Delta V \) represents the changes in water body storage. The amount of water entering the Urmia Lake is subject to the following parameters (Jensen, 2010):

\[ Q_{\text{input}} = f\left( Q_{\text{input}}, R_{\text{input}}, R_{\text{input}}', G_{\text{input}} \right) \]  

(3)

where \( P \) is the direct precipitation on the lake, \( R_{\text{input}} \) is the amount of water entering the lake by permanent rivers, \( R_{\text{input}}' \) is runoff from the land basin, and \( G_{\text{input}} \) is groundwater inflow. Since the Urmia Lake is a closed lake, the amount of water released from the Urmia Lake is also subject to the following parameters (Jensen, 2010):

\[ Q_{\text{output}} = f\left( E, G_{\text{output}} \right) \]  

(4)

where \( E \) is evaporation from the water surface, and \( G_{\text{output}} \) is groundwater outflow (Jensen, 2010). If \( Q_{\text{input}} = G_{\text{output}} \), by inserting Equations 3 and 4 in Equation 2:

\[ R_{\text{input}}' = \Delta V - (P + R_{\text{input}} - E) \]  

(5)

by plotting the graph of Equation 5:

Figure 4. Changes in the amount of runoff entering the Urmia Lake during 1999-2006

The figure above shows that since 2002, the amount of temporary runoff into the lake has been increasing, and then the amount of this runoff decreases and increases again, which is consistent with the SC graph in Figure 3. As a result, this can indicate the relationship between the water level of the lake and the SC in the catchment area, and the temporary increase in the Urmia Lake’s water level during 2002 and 2005 can be due to temporary runoff from snow melting. For this reason, the purpose of this paper is the modeling and analyzing of the relationship between hydrological data and SC with Urmia Lake’s water level. Since the datasets are non-linear in the real-world application therefore, the linear regression model can fail to generalize. Thus developing a nonlinear relationship between a pair of input and output is necessary (Pasupa et al., 2016). In this research, in order to accommodate the non-linearity, a new strategy for water level estimation of Urmia Lake based on SVM regression with various kernel types has been proposed. In the proposed method, four scenarios are considered with different variables as inputs of the SVR including SC in the whole catchment area of the Urmia Lake, direct precipitation and direct evaporation of the lake’s surface. As demonstrated in Figure 5, it is worth noting that all of the scenarios resulted in the water level of the lake (Figure 5).

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-4-W18-219-2019 | © Authors 2019. CC BY 4.0 License.
### 3.1 Support Vector Regression (SVR)

In most cases, the researcher had attempted to establish a linear relationship between the input data and the corresponding target data. However, with the discovery of nonlinearity in the nature of the data, the focus has shifted towards the nonlinear prediction. Although there are many kinds of nonlinear statistics used in the literature for the water level forecasting, most of them require the nonlinear model to be specified before the estimation is done. In this paper, since the water level, climatic and hydrological data are nonlinear and follow a very irregular trend, SVR has evolved out to be a better technique to bring out the structural relationship between the various entities (Abhishek et al., 2012). The theoretical foundation of the SVM has been developed by (Vapnik, 2013). SVM can also be used as a regression method, maintaining all the main features that characterize the algorithm (maximal margin) (Sayad, 2019).

There are two linear and nonlinear methods for SVR. Training an SVR with a linear kernel is faster than with other kernels, and it needs fewer parameters to optimize. The linear SVR is easy to interpret but may have low predictive accuracy. On the other hand, the non-linear SVR is more difficult to interpret but can be more accurate. In SVR, the kernel functions transform the data into a higher dimensional feature space to make it possible to perform the linear separation (Figure 6) (Nikraftar et al., 2015), (Boueshagh et al., 2018), (Sayad, 2019).

![Nonlinear SVR method](Figure 6. Nonlinear SVR method (Sayad, 2019))

Equations 6 and 7 represent the Gaussian and polynomial types of non-linear kernels, respectively (Sayad, 2019).

\[
\text{Polynomial} : K(x_i, x_j) = (x_i \cdot x_j)^d \\
\text{RBF} : K(x_i, x_j) = \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right)
\]

In this study, the Regression Learner App is used to carry out the analysis of the input data. This app implements linear epsilon-insensitive SVM regression, which ignores prediction errors that are less than some fixed number \(\epsilon\). The support vectors are the data points that have errors larger than \(\epsilon\). In this paper, Quadratic (Q), Cubic (C) and Gaussian or Radial Basis Function (RBF) kernels are used to train the model, and also Box constraint (BC), Epsilon (E) and Kernel Scale (KS) parameters have been tuned. For each of the scenarios, 70% of the data is considered as training data and 30% as testing data randomly, and the testing data is not included in the calculations. In addition, the information from the different inputs was scaled to \([0, 1]\) range and also standardized.

For each of the aforementioned scenarios through trial and error, the SVR parameters are shown in Table 1:

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Parameters</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC</td>
<td>---</td>
<td>0.572</td>
<td>---</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>---</td>
<td>0.068</td>
<td>---</td>
<td>0.069</td>
<td></td>
</tr>
<tr>
<td>KS</td>
<td>1.03</td>
<td>---</td>
<td>1.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BC</td>
<td>---</td>
<td>0.33</td>
<td>---</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>---</td>
<td>0.033</td>
<td>---</td>
<td>0.071</td>
<td></td>
</tr>
<tr>
<td>KS</td>
<td>---</td>
<td>1.09</td>
<td>---</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>RBF</td>
<td>BC</td>
<td>0.53</td>
<td>1.91</td>
<td>2</td>
<td>1.73</td>
</tr>
<tr>
<td>E</td>
<td>0.04</td>
<td>0.06</td>
<td>0.11</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>KS</td>
<td>0.1</td>
<td>1.09</td>
<td>1.1</td>
<td>0.92</td>
<td></td>
</tr>
</tbody>
</table>

### 4. RESULTS

The test criteria for the performance measures of the forecasting methods used in this study are the root mean square error (RMSE) and R-squared (R\(^2\)). After the water level modeling, these performance measures are outlined in Table 2.

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Kernel</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>In: PRE</td>
<td>RBF</td>
<td>87%</td>
<td>80.26%</td>
</tr>
<tr>
<td>Out: WL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In: PRE &amp; EVA</td>
<td>Q</td>
<td>75%</td>
<td>77.93%</td>
</tr>
<tr>
<td>Out: WL</td>
<td>C</td>
<td>90%</td>
<td>87.39%</td>
</tr>
<tr>
<td>In: SC &amp; PRE</td>
<td>RBF</td>
<td>89%</td>
<td>87.57%</td>
</tr>
<tr>
<td>Out: WL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In: SC &amp; PRE &amp; EVA</td>
<td>Q</td>
<td>90%</td>
<td>92.13%</td>
</tr>
<tr>
<td>Out: WL</td>
<td>C</td>
<td>94%</td>
<td>93.41%</td>
</tr>
<tr>
<td>RBF</td>
<td>97%</td>
<td>96.97%</td>
<td>0.08</td>
</tr>
</tbody>
</table>

According to the results, the application of the SVR has been successfully demonstrated in this paper for mean monthly Urmia Lake water level prediction. Results generally prove that for this case study the RBF SVR is a superior model to the polynomial and produced the most accurate results for modeling the water level behavior. Furthermore, in the second scenario, similar results were obtained for the Cubic and RBF kernels,
which show that the degree-3 polynomial like the RBF kernel is capable of modeling the relationship between precipitation and evaporation with the water level. Based on the third and the fourth scenarios, the impact of snow on the quality of modeling is also observable. In Figures 7, 8, 9, 10, and 11 the outputs of the SVR for all the scenarios are shown and compared with the actual water level. The results indicate a high correlation between the calculated values and the actual values of the Urmia Lake water level.

\[ R^2 = 87\% \] (a)

\[ R^2 = 75\% \] (b)

\[ R^2 = 90\% \] (c)

\[ R^2 = 89\% \] (d)

\[ R^2 = 95\% \] (e)

\[ R^2 = 90\% \] (f)

\[ R^2 = 94\% \] (g)
According to Figures 7 and 8, and the $R^2$ and RMSE metrics' value presented in Table 2, it can be concluded that the fourth scenario and the RBF kernel, which have the best results, can be used to establish a relationship between snow and other hydrological parameters with the water level of the Urmia Lake. It is also determined that the direct precipitation in the lake, the SC in the catchment area and the direct evaporation of the lake surface have the highest impact on the water level of the Urmia Lake, respectively. The results of the validation test of the estimating model shown in Table 2 indicate that the $R^2$ value for the fourth scenario (with all kinds of kernel functions especially the RBF SVR) were more closely aligned to the $R^2$ value of the training dataset than the other scenarios. On the other hand, the $R^2$ value for the first scenario was more different for the training and testing data than for the other scenarios, which means that the direct precipitation alone is not capable of modeling the Urmia Lake’s water level.

**Figure 7.** Correlation between retrieved and observed water level for training data: (a) RBF kernel, first scenario; (b) Q kernel, second scenario; (c) C kernel, second scenario; (d) RBF kernel, second scenario; (e) RBF kernel, third scenario; (f) Q kernel, fourth scenario; (g) C kernel, fourth scenario; (h) RBF kernel, fourth scenario.

**Figure 8.** Correlation between retrieved and observed water level for testing data: (a) RBF kernel, first scenario; (b) Q kernel, second scenario; (c) C kernel, second scenario; (d) RBF kernel, second scenario; (e) RBF kernel, third scenario; (f) Q kernel, fourth scenario; (g) C kernel, fourth scenario; (h) RBF kernel, fourth scenario.

**Figure 9.** Comparisons of retrieved and observed water level for testing data: (a) RBF kernel, first scenario; (b) Q kernel, second scenario; (c) C kernel, second scenario; (d) RBF kernel, second scenario; (e) RBF kernel, third scenario; (f) Q kernel, fourth scenario; (g) C kernel, fourth scenario; (h) RBF kernel, fourth scenario.

$R^2 = 97\%$
In this paper, the water level estimation of Urmia Lake was done using the SVR method. Analyzing models in various scenarios indicated that this method has a high performance, especially in the fourth scenario because the number of the predictors in this scenario is more than other scenarios, and when the number of predictors is higher, the SVR performs better. Also, the presence of the snow parameter in this scenario is another reason for more accurate modeling in the fourth scenario. Moreover, the outputs of various kernels showed that due to the complexity of the connections between the inputs and the output, the RBF kernel is better than other kernels. The results showed the significant impact of SC on retrieving the water level of the Urmia Lake at the desired period, and the high performance of SVR method in estimating the water level using hydrological parameters of the basin. The modeling results indicate that the direct precipitation in the lake, the SC in the catchment area and the direct evaporation of the lake surface have the highest impact on the water level of the Urmia Lake, respectively. As a result, because of the relationship between the water level of the Urmia Lake and the snow parameter in desired period (2000-2005), and besides, according to the water budget procedure, the temporary increase in the water level of the Urmia Lake during 2002-2005 may be due to the temporary runoff from snow melting. Also, if the amount of water entering the lake is greater than the amount of the evaporation, or at any time, the water entering the lake is not less than the amount of the evaporation, the lake will have no shrinkage. Therefore with any rainy year and the snowmelt runoff, the Urmia Lake will naturally be recovered.

REFERENCES


5. CONCLUSION

Figure 9 represents the visual comparison of the testing results with the measured water level of Urmia Lake. The minor differences between the testing results and the measured water level show a relatively high quality of modeling for every scenario. Figure 9 shows that the first and the second scenarios are not able to model the lower water levels while modeling the upper water levels well, and then, in the third and the fourth scenarios, with the advent of the snow parameter, the accuracy of the modeling process is increased for lower water levels.

In Figures 10 and 11, the yellow, red, and the blue colors indicate the RBF, Cubic, and Quadratic kernels, respectively. As the numeric analysis, scenario 4, as well as the RBF kernel, has the highest $R^2$ and the lowest RMSE.

Based on the water budget procedure, from the end of 2002 to the beginning of 2005, the amount of $R_{\text{input}}$ (runoff from snow melting) has increased. Additionally, according to Figure 3, at the end of 2001, a large peak is visible in the SC time series. It can be said that the presence of snow peaks at the end of 2001 has controlled the downward trend of Urmia Lake’s water level and then the occurrence of rainy years causes a temporary increase in the lake’s water level over the desired period (2002-2005).


