STATISTICAL ANALYSES AND HYDROMETEOROLOGICAL PARAMETERS OF LONG-YEAR WATER LEVELS AND VOLUMES: EXAMPLE OF LAKE EGIRDIR

Atilgan Atilgan[1], Ali Yucel[2], Monika Markovic[3]
[1]Alanya Alaaddin Keykubat University, Turkey; [2]Osmaniye Korkut Ata University, Turkey;
[3]Josip Juraj Strossmayer University of Osijek, Croatia
atilganatilgan@isparta.edu.tr, aliyucel965@gmail.com, monika.markovic@fazos.hr

Abstract. Global warming, climate change, rapid urbanization and industrialization, population growth, unplanned and random water use considerably affect the lake water level (*LWL*) and lake water volume (*LWV*), which is very important since the lakes are clean water sources. For this study, Lake Egirdir, with a tectonic structure and basin area of approximately 3445.6 km^2 within the borders of Isparta province, was chosen. In the study, long-year lake water level measurement station values of Lake Egirdir and long-year temperature, precipitation, relative humidity, wind speed values of the Egirdir meteorology station were used as variables. To develop equations between the Egirdir lake water level, volume and hydro-meteorological variables, the best fitted sub-equation models analysis, analysis of variance and multiple regression analyses (MRA) were performed. Their statistical relevance was examined with 5% significance level F, correlation (r) and probability (p) tests. Analysis of variance of each MRA equation developed to determine *LWL* and *LWV* estimation values was evaluated. In line with these values, it has been determined that the MLRA equation is more suitable than the Multiple Nonlinear Regression Analysis (MNRA) equation in estimating *LWL*, and the MNRA equation is much more suitable than the Multiple Linear Regression Analysis (MLRA) equation in estimating *LWV*. With the help of the equations developed for *LWL* and *LWV* estimations, it has been concluded that the Egirdir lake basin water management planning and project design studies within the hydrological cycle can have an idea beforehand. As a result, we think that it is very important to carry out the necessary planning and management studies in advance by revealing the current status of water resources in the future by carrying out these and similar studies.

Keywords: Egirdir, hydrometeorological parameters, regression analysis.

Introduction

The need for freshwater is continuously increasing due to the gradual decrease in the amount of usable fresh water in the world and the increase in population. In this case, water requires good management of resources [1; 2]. Modeling studies are also the most important method used in the management of these resources. During lake management, the effects of many planned studies can be determined with the help of modeling [2]. Ensuring the management of water, it can be realized sustainably by evaluating the hydrological, ecological, and economic aspects together and in harmony. With the last quarter of the twentieth century, water management at the basin scale has come to the fore, and the current situation is revealed with a holistic approach and solutions are sought for problems [3]. As the main reasons for the changes in the decrease in the levels of the lakes, researchers stated various factors such as changes in ground cover, changes in land use, urbanization, increase in agricultural and animal water needs, excessive use of the resources that feed the lakes and dam construction [4]. Researchers stated that besides human factors, climatic factors also cause changes in lake water levels [4; 5]. Lakes have human attention for centuries, both in terms of fresh water supply, which is inevitable for life, as well as important resources for many practices. From the perspective of the history, many important civilizations were established and rooted in areas where freshwater resources were found. However, especially in the last two centuries, when industrialization took place, the excessive proliferation of human beings and their insensitivity to the ecosystem have brought many environmental problems. These problems have gained momentum in the last century and have reached enormous dimensions [6].

In the study, Egirdir Lake water level and water volume modeling was performed using the multiple regression analysis methods, using water balance components and water level records. With the help of the created regression model, possible extreme annual water levels in the future were tried to be estimated.

Material and methods

Lake Egirdir is an important freshwater lake located in the Lakes region in the Western Mediterranean Region of Turkey. Administratively, it is located within the borders of Isparta province between 35º37’41”-38º16’55” North latitudes and 30º44’39”-30º57’43” East longitudes. The average height of the lake is 917.7 m, its average area is 481.5 km^2 and its water volume is 4.44 billion m^3 [7]. In the study, *LWL* and *LWV* values measured by DSI, during 1988-2018 period were used. Long-term (1988-2018) temperature, precipitation, relative humidity, and wind speed values obtained from the general directorate of meteorology were used as materials [8; 9].

Autocorrelation analysis is a statistical method used to determine the correlation (internal dependence) between consecutive values in the values of the observation series ordered in time or locality [10-14]. Autocorrelation coefficient (r_k):

$$r_k = \frac{\sum_{i=1}^{N-k}\left[(x_i - \bar{x}_i)(x_{i+k} - \bar{x}_{i+k})\right]}{\sum_{i=1}^{N}(x_i - \bar{x}_i)^2}, \tag{1}$$

where x_i – i data of the observation series,
$\bar{x}$ – mean of the observation series,
i – ordinal number of values in the observation series,
k – number of lags (k = 1, 2, ..., N),
N – total number of observations in the observation series.

The statistical significance of the calculated autocorrelation coefficient is made with the t test at the chosen significance level (α). Accordingly, the t test statistic (t_k) is calculated as follows:

$$t_k = \frac{r_k}{SE(r_k)}. \tag{2}$$

The standard error of each autocorrelation coefficient is calculated for $k = 1$ and $k \geq 2$ as follows:

$$\text{for } k = 1 \;\; SE(r_1) = \sqrt{1/N} \;\; \text{for } k \geq 2 \;\; SE(r_k) = \sqrt{\left(1 + 2 \cdot \sum_{k=1}^{N-1} r_k^2\right)/N},$$

where $SE(r_k)$ – standard error of lag number,
r_k – autocorrelation coefficient,
k – lag number (k = 1, 2, …, N),
N – total number of observations in the observation series.

The t_k values calculated for each lag number (k = 1, 2, 3,… N) in the observation series are compared with the selected α significance level and the t distribution table value (t_t) at n-1 degrees of freedom. If $t_k < t_t$, it is decided that there is no autocorrelation (r_k) between the values of the observation series, and if $t_k > t_t$, there is autocorrelation [17]. The correlogram is used for the visual (graphical) representation of the statistical significance of each calculated r$_k$. Each r_k value in the correlogram must be between the calculated upper confidence limit $UCL(r_k) = t_t \cdot SE(r_k)$ and lower confidence limit $LCL(r_k) = t_t \cdot SE(r_k)$ values [10; 12; 14-17].

Multiple Regression Analysis. In statistical studies, there may be a linear or non-linear relationship between dependent and independent variables. This relationship is determined by the correlation analysis to be made between the dependent and independent variables. In this analysis, the direction and size of the relationship between the variables are determined as well as the type of mathematical function. According to the type of mathematical function, they are divided into linear and nonlinear regression models [13; 17].

Multiple Linear Regression Analysis is a model used for linear modeling of the relationship between the dependent variable and more than one independent variable [13; 17-20].

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n + \varepsilon, \tag{3}$$

Multiple Nonlinear Regression Analysis is a model used to minimize the estimation error (ε) when the relationship between more than one independent variable and the dependent variable is not linear.

$$Y = \alpha + X_1^{\beta_1} X_2^{\beta_2} X_3^{\beta_3} \ldots X_n^{\beta_n} + \varepsilon, \tag{4}$$

where $\alpha, \beta_1, \beta_2, .., \beta_n$, – coefficients of equations,
Y – dependent variable,

X_1, X_2, .., X_n – independent variables,
ε – error term [17-21].

Best Sub-Equation Group Analysis. Subgroups of equations are formed by using all independent variables that have an effect on the dependent variable. It is a statisti7777cal method used to determine the best sub-equation group containing the least independent variable among all sub-equation groups created. For this, all subgroups of equations are ordered according to their regression coefficients (R^2). The best subgroup of equations among them is determined by the F test.

$$F = \frac{\left(R_p^2 - R_{p-i}^2\right)(p-i)}{\left(1-R_p^2\right)/(N-p-1)}, \tag{5}$$

where R_{p-i} – regression coefficient of the previous subgroup of equations,
R_p – regression coefficient of the next subgroup of equations,
p – number of variables used in the previous sub–equation group,
i – number of variables used in the next sub–equation group,
N – total number of observations [13; 17].

The calculated F values are compared with the F_t value of the F distribution ($F\alpha$, (N-i), (N-p-1)). If $F > F_t$, this process is continued among the other sub-equation groups that follow. However, when $F < F_t$, the test is stopped, and it is decided that the previous sub-equation group is statistically the best sub-equation group [18; 22; 23].

Analysis of Variance is used to analyze how the independent variables affecting the dependent variable in the developed estimation equation interact with each other and the effects of these interactions on the dependent variable. Statistical significance is determined by comparing it with the F test at the chosen significance level (α) [13; 17; 21].

Results and Discussion

The changes in the long-year *LWL* and *LWV* values of Lake Egirdir were investigated. While the long annual *LWL* average of Lake Egirdir was 917.095 m, it was determined that it reached the minimum water level with 916.028 m in 1995 and the maximum water level with 918.051 m in 2004. When *LWV* values were examined, it was determined that while the average was 3377.488 hm^3, it reached a minimum value of 2998.937 hm^3 in 2009 and a maximum value of 3774.435 hm^3 in 2000. In general, in Figure 1, it has been determined that the differences in the change between *LWL* and *LWV* values of the lake have increased since about the middle of 1999, while it was determined that this difference was more between the years 1999-2018.

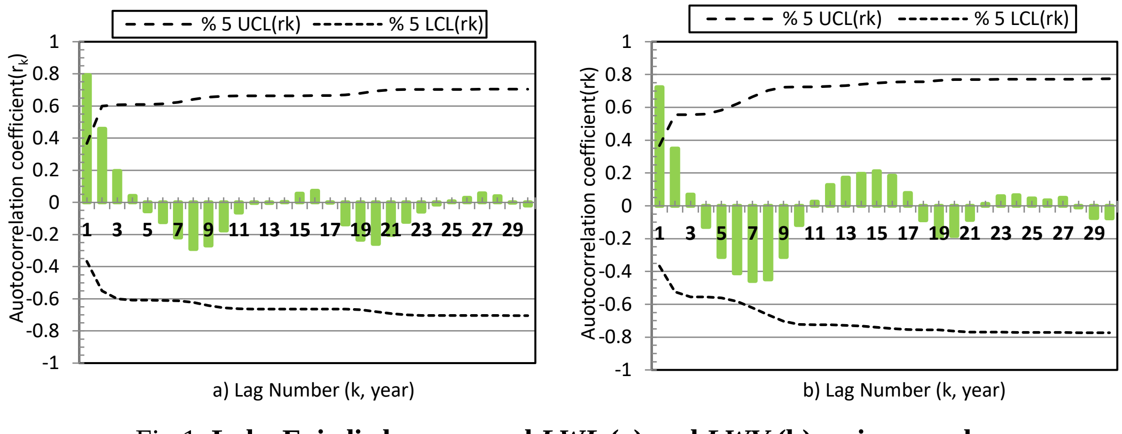

Fig 1. **Lake Egirdir long-annual *LWL* (a) and *LWV* (b) series correlograms**

We can say that the demand for water volume from Lake Egirdir increased more after the middle of 1999. Therefore, it has been emphasized that *LWL* and *LWV* values may be addictive. For this purpose, the serial dependencies of the lake to *LWL* and *LWV* values between 1988 and 2018 were analyzed by autocorrelation analysis and their correlograms were created (Fig. 1). When the correlograms of *LWL* and *LWV* series in Lake Egirdir are examined on an annual basis, it is seen that the k = 1 delayed (year) serial dependence (r_1 = 0.792 in *LWL* and r_1 = 0.724 in *LWV*) is very important, and that *LWL* and *LWV* values compared to *LWL*1 and *LWV*1 values 1 year ago have been determined to be involved (Fig 1). In other words, it was determined that there is a very significant serial dependence between *LWL* and *LWV* values and 1 year later of *LWL*1 and *LWV*1 values. Fig 2 has been prepared in order to determine whether the current year's lake water level and water volume are effective on the previous year's values due to the continuity of the lake's water circulation.

When Figure 2 is examined, it has been decided that the changes between *LWL* and *LWL*1 and *LWV* and *LWV*1 values are parallel to each other, therefore, *LWL*1 and *LWV*1 values are effective independent variables in modelling *LWL* and *LWV* values. The correlation (*r*) analyses of the measured *LWL* and *LWV* values (dependent variable) and *LWL*1, *LWV*1, temperature, precipitation, proportional humidity, and wind speed values (independent variables) were performed separately both to the measured values and to the logarithmically transformed values. The obtained results were compared with r and p values at 5% significance level (< 0.05). It was determined that *LWL* was affected by *LWV*, *LWL*1, *LWV*1, annual minimum proportional humidity, annual maximum wind speed variables, and *LWV* by *LWV*, *LWV*1, *LWL*, annual maximum precipitation, annual maximum proportional humidity variables, respectively. Other independent variables were determined to be insignificant.

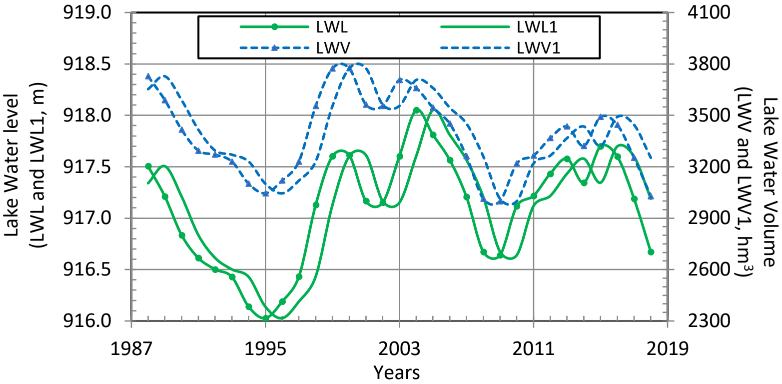

Fig 2. **Changes between *LWL* and previous year's *LWL*1, *LWV* and previous year's *LWV*1 values**

With the analysis of the best sub-equation models, the best sub-equation models that contain very few independent variables that are effective on *LWL* and *LWV* dependent variables determined for multiple regression analyses were determined. According to the analysis results, respectively; *LWV*, *LWV*1, *LWV*, annual average precipitation, annual minimum wind speed variables were effective on *LWL*. *LWL*, *LWH*1, *LWL*1, annual average precipitation and annual minimum wind speed variables were found to be effective for *LWH*. According to the variance analysis results of each MRA equation developed for the estimation of *LWL* and *LWV* values it has been determined that the MLRA equation (*FH* = 323.62) is more suitable than the MNRA equation (*F* = 281.92) in *LWL* estimation. It has been determined that the MNRA equation (*F* = 285.32) is much more suitable than the MLRA equation (*F* = 281.60) in *LWV* estimation (Table 1). The average and standard deviations of the long-year P_{ave}, RH_{ave} and RH_{min} variables of Lake Eğirdir are calculated as 7.280 ± 1.244 mm·year^{-1}, 66.934 ± 3.195% and 35.174 ± 5.254%.

Table 1

Developed MLRA and MNRA equations and statistical properties

Models	Developed Multiple Linear Regression Equation (MLRA) and Multiple Nonlinear Regression Equation (MNRA)	Correlation coefficient (*r*)	Standard deviation (*s*)
MLRA	$\boldsymbol{LWL = 82.0 + 0.9103 \cdot LWL_1 + 0.002258 \cdot LWV - 0.002043 \cdot LWV_1 - 0.01339 \cdot RH_{min}}$	0.9915	0.0809972
	$LWV = -29220.0 + 0.9060 \cdot LWV_1 + 411.7 \cdot LWL - 379.7 \cdot LWL_1 + 5.86 \cdot RH_{min} - 33.1 \cdot V_{min}$	0.9913	33.4717
MNRA	$LWL = 0.242 \cdot LWL_1^{0.9178} \cdot LWV^{0.008364} \cdot LWV_1^{-0.007607} \cdot P_{ave}^{-0.000182} \cdot RH_{min}^{-0.000485}$	0.9913	1.000085
	$\boldsymbol{LWV = -23.8 \cdot LWV_1^{0.9126} \cdot LWL^{113.27} \cdot LWL_1^{-105.16} \cdot P_{ave}^{0.0236} \cdot RH_{min}^{0.0568}}$	0.9914	1.00991

In order to determine the most suitable equation model for *LWL* and *LWV*, minimum correlation coefficient (r) and standard deviation (s) values were taken into account. It has been determined that the appropriate model for *LWL* ($r = 0.9915$, $s = 0.0809972$) is MLRA, and for *LWV* ($r = 0.9914$, $s = 1.00991$) the MNRA models are suitable. With the help of the equations developed for *LWL* and *LWV* estimations, it has been concluded that it is possible to make predictions in the water management planning and projecting studies of the Lake Egirdir basin within the hydrological cycle. Göncü et al. [4] stated that the effect of climatic factors and anthropogenic effects should be separated from each other in the trend analysis studies they conducted with the lake water level. In this way, they concluded that the trends in the lake level changes could be better understood. Our study is similar to this study. Batur et al. [24] in their study of Lake Van stated that the water level in the lake directly depends on the precipitation-flow-evaporation interaction prevailing in the basin and the regime of hydrometeorological variables. This study provides parallelism with our study.

Conclusions

In the study, long-year meteorological values were used together with the measured *LWL* and *LWV* values of Lake Egirdir. Due to the continuity of water circulation in the natural structure of the lake, it was determined whether the current year's *LWL* was effective on the previous year's *LWL*1, and the current year's *LWV* was effective on the previous year's *LWV*1. It was decided that the changes between *LWL* and *LWL*1 and *LWV* and *LWV*1 values were parallel to each other, and therefore *LWL*1 and *LWV*1 values were effective independent variables in modeling *LWL* and *LWV* values. In addition, variance analyses of each MRA equation developed to determine *LWL* and *LWV* estimation values were evaluated. In line with these values, it has been determined that the MLRA equation is more suitable than the MNRA equation in estimating *LWL*, and the MNRA equation is much more suitable than the MLRA equation in estimating *LWV*. With the help of the equations developed for *LWL* and *LWV* estimations, it has been concluded that the Egirdir lake basin water management planning and project design studies within the hydrological cycle can have an idea beforehand. In this context, we believe that there will be more studies that will reveal the status of water resources. As a result, we think that it is very important to carry out the necessary planning and management studies in advance by revealing the status of water resources in the future by carrying out these and similar studies.

References

[1] Cooke G.D., Lombardo P. Brant C. Shallow and deep lakes: Determining successful management options. LakeLine (NALMS), 2001, 21, pp. 42-46.

[2] Mercan D.E., Kabdaşlı S. Hydrodynamic modeling of Beysehir Lake for 1998 water year itüdergisi/d mühendislik 2008, 7(2), pp. 75-82 April 2008. (In Turkish)

[3] Uğurluoğlu A. Su Havzalarında Hidrolojik Modelleme: Manyas Gölü Yukarı Havzası Uygulaması, Orman ve Su İşleri Bakanlığı, Uzmanlık tezi 2017. (In Turkish)

[4] Göncü S., Albek E.A., Albek M. Trend Analysis of Burdur, Eğirdir, Sapanca and Tuz Lake Water Levels Using Nonparametric Statistical Methods, AKU J. Sci. Eng.17, 2017 025401, pp. 555-570.

[5] Yıldırım, Ü., Erdoğan, S. Uysal, M. Changes in the Coastline and Water Level of the Akşehir and Eber Lakes Between 1975 and 2009. Water Resources Management 2011, 25(3), pp. 941-962.

[6] Anonymous. Göller ve Sulak Alanlar Eylem Planı 2017-2023, T. C. Orman ve Su İşleri Bakanlığı, Su Yönetimi Genel Müdürlüğü, 2017, 194 sayfa, Ankara. (In Turkish)

[7] Kargı H. Gözbebeğimiz Eğirdir Gölü Paneli: Sorunlar & Çözümler Etkinlik Kitabı, 6 Şubat 2019, Süleyman Demirel Üniversitesi, 2019, pp. 51-55 sayfa. (In Turkish)

[8] DSİ. Eğirdir Gölü Ölçüm İstasyonu 1988-2018 Yılları Su Seviyesi Verileri, Devlet Su İşleri 18. Bölge Müdürlüğü, 2019, Isparta. (In Turkish)

[9] DMİ. Isparta-Eğirdir İlçesi 1988-2019 Yılları Meteorolojik Verileri, 2019, Ankara. (In Turkish)

[10] Anderson R. L. Distribution of The Serial Correlation Coefficient. Annals of Mathematical Statistics, 1942, 13(1), pp. 1-13.

[11] Matalas N. C. Autocorrelation of The Rainfall and Streamflow Minimums, Statistical Studies in Hydrology, Geological Survey Professional Paper 434-B, 1963, 14 pp., Washington, D. C., USA.

[12] Kottegoda N.T. Stochastic Water Resources Technology, 1st Edition, The MacMillan Press Ltd., 1980, 384 p., Hong Kong.

[13] Pedhazur E. J. Multiple Regression in Behavioral Research Explanation and Prediction, Third Edition, Wardsworth &Thomson Learning, Inc., 1997, 1072 p., Florida-USA.
[14] Chatfield C. Time-Series Forecasting, 2000, 265 p., Chapman & Hall/CRC, USA.
[15] Olofintoye O., Adeyemo J., Otieno F. Impact of Regional Climate Change on Freshwater Resources and Operation of the Vanderkloof Dam System in South Africa, Global Warming-Impact and Future Perspective, 2012, DOI: 10.5772/50414, pp. 165-184.
[16] Box G. E. P., Jenkins G. M. Reinsel G. C. Time Series Analysis Forecasting and Control, 3rd Edition, Prentice-Hall International Inc., 1994, 614 p., New Jersey, USA.
[17] Chatterjee S. Simonoff J. S. Handbook of Regression Analysis, A John Wiley & Sons, Inc., Publication, 2013, 236 p., USA.
[18] Berry W. D. Feldman S. Multiple Regression in Practice, A SAGE University Paper, Quantitative Applications in The Social Sciences Series No: 50, SAGE Publications, Inc., 1685, 95 p., California.
[19] Özdamar K., SPSS ile Biyoistatistik, Kaan Kitabevi, 2003, Eskişehir. (In Turkish)
[20] Güriş S., Astar M. Bilimsel Araştırmalarda SPSS ile İstatistik. 1. Baskı, İstanbul, DER Kitabevi, 2019, 512 sayfa. (In Turkish)
[21] Maity R. Statistical Methods in Hydrology and Hydroclimatology, Springer Transactions in Civil and Environmental Engineering Book series. Civil Engineering, 2018, 366 p, Singapore.
[22] Draper N. R., Smith H. Applied Regression Analysis, Wiley Series in Probability and Statistics: Probability and Statistics Section Series, John Wiley and Sons Inc., 1981, 709 p., New York, USA.
[23] Kleinbaum D. G., Kupper L. L., Muller K. E. Nizam A. Applied Regression Analysis and Other Multivariable Methods, 3rd Edition, An International Thomson Publishing Company, 1998, 792 p., California, USA.
[24] Batur E., Kadıoğlu M., Özkaya M., Sağban M., Akın İ. Water level modelling of Lake Van, Water Resources, 2009, 2, pp. 27-40. (In Turkish)