GALAXY INTERNATIONAL INTERDISCIPLINARY RESEARCH JOURNAL (GIIRJ)
ISSN (E): 2347-6915
Vol. 9, Issue 12, Dec. (2021)

FORECASTING GROUNDWATER EVAPORATION USING MULTIPLE LINEAR REGRESSION

Yaxshiboyev R.E,

Gaybullayev E.E,

Husanov U.A,

Ochilov T.D

ABSTRACT

Models of regression analysis and classification of time-series data based on machine learning algorithms allow solving the problem of forecasting the state of the region in various fields, including agriculture. One of the problems in this area is soil salinity, one of the main causes of salinization being associated with rising groundwater levels. This paper is devoted to defining a model for predicting groundwater evaporation using a multiple variable linear regression method using geographic data from the region. Data from of Khorezm region between 1980 and 2010 were used as input data for the construction of the model, and a training sample was developed based on this data. A correlation analysis was performed to study the relationship between the sample variables, and a three-variable linear regression model consisting of precipitation, water evaporation, and air temperature was used to predict the groundwater level and to increase the accuracy of the model. The method of clearing the data in the training sample from interference is also presented in this article.

Keywords: analysis, water evaporation, Prediction, linear regression, geoinformation data, OLS

INTRODUCTION

Groundwater information processing is a complex task. Assessment of groundwater status is inextricably linked to many ecosystems, such as charging, salinization, shallow water levels, and surface water events [1]. Because the resource is underground, tracking or measuring it is a complex process. The most common way to estimate a resource is to look at the groundwater level measured from these wells. For water resources to use and practice artificial intelligence can produce excellent results.

Today in the world the demand for drinking water is growing and this arises the problem of water shortage. Higher temperatures and little rainfall leads to drought. To solve this problem, modern technologies, programs, mathematical models and algorithms are being developed. This paper is provides the results of research of constructing a model for predicting groundwater elevation using a multiple variable linear regression method using geo-spatial data from the region.

MATERIAL AND METHODS

Spatial data form the basis of information support for geographic information systems. Modern analysis of geospatial data allows to combine a geographic information system with business intelligence, which leads to high-quality, fast decision-making by reducing the time for searching and analyzing the necessary information. Spatial analysis allows the map to be used as one of the standard measurements, such as time. [1,9]

Linear regression is one of the most important and widely used regression techniques. This is the simplest regression method. One of its advantages is the ease of interpretation of the results.

Linear regression of some dependent variable y on a set of independent variables $x = (x_1,\ldots, x_r)$, where r is the number of predictors, assumes that a linear relationship between y and x: $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r + \varepsilon$. This is a regression equation. [2,10]

β_0, β_1,..., are regression coefficients, and ε is a random error.

Linear regression calculates the estimation functions of the regression coefficients, or simply predicted weights of the measurement, denoted as $b_0, b_1,\ldots, b_r$.

They define the estimated regression function $f(x) = b_0 + b_1 x_1 + \cdots + b_r x_r$. This function captures the dependencies between inputs and outputs quite well. [3,11]

For each observation result i = 1, ..., n, the estimated or predicted answer $f(x_i)$ should be as close as possible to the corresponding actual answer y_i.

The differences $y_i - f(x_i)$ for all observations are called residuals. Regression determines the best predicted measurement weights that correspond to the smallest residuals. [4,12]

To get the best weights, you need to minimize the sum of squares residual (SSR) for all observations:

$SSR = \Sigma_i (y_i - f(x_i))^2$.

This approach is called the least squares method.

Python packages for linear regression:

NumPy is a fundamental scientific package for fast operations on one-dimensional and multidimensional arrays. NumPy eases the math routine and is of course open source. [5,13]

The scikit-learn package is a library widely used in machine learning. scikit-learn provides values for preprocessing data, downsizing, implements regression, classification, clustering, etc. It is open-source, just like NumPy.

You can provide several optional parameters to the LinearRegression class:

• fit_intercept is a boolean (True by default) parameter that decides whether to calculate the segment b_0 (True) or treat it as equal to zero (False).

• normalize is a boolean (False by default) parameter that decides whether to normalize the input variables (True) or not (False).

• copy_X is a boolean (True by default) parameter that decides whether to copy (True) or overwrite input variables (False).

• n_jobs is an integer or None (default) representing the number of processes involved in parallel computing. None means no processes, -1 uses all available processors. [6]

RESULTS

A geographical map of the Republic of Uzbekistan was chosen (Fig. 1.).

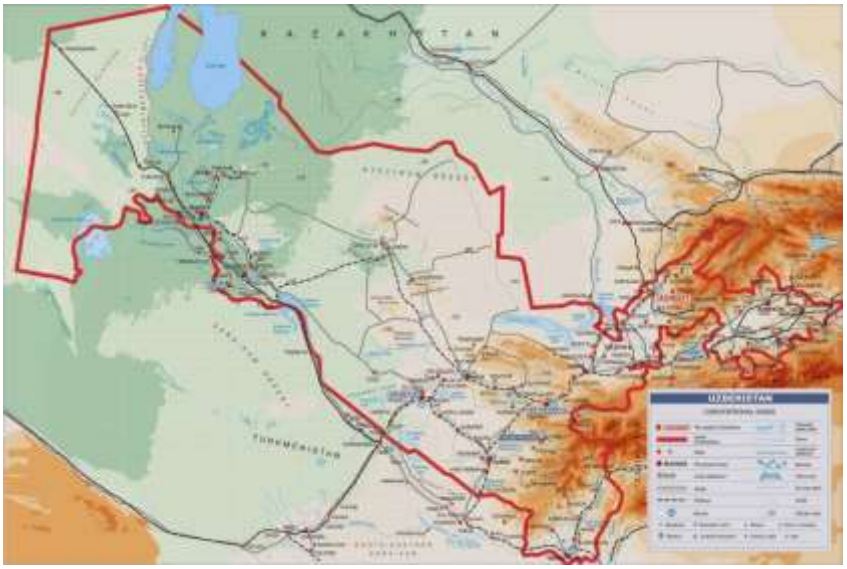

Fig. 1. Geographic map of Uzbekistan

Based on this map, a Data set was prepared about the nature, air temperature and relative humidity of the selected object. Based on the created mathematical model, matrix and algorithm, predictions are performed in the Python programming language. All data in the Data set is analyzed and compared with each other. [15]

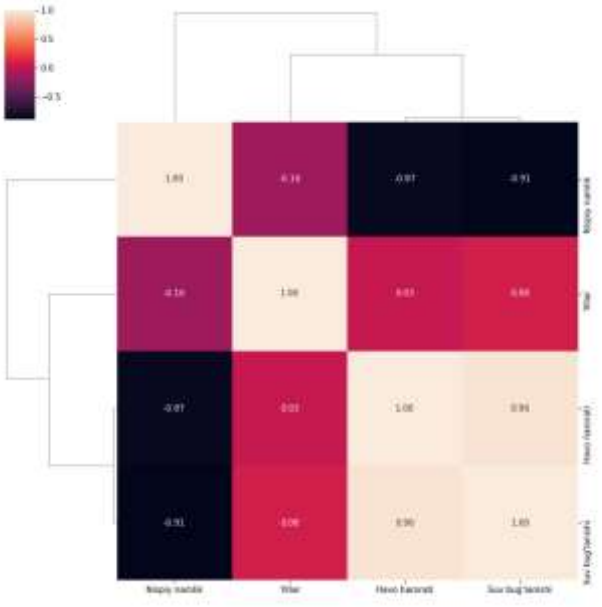

Fig.2. Correlation matrix

The data is analyzed using the Data set for the selected mathematical model and correlation matrix. Then the analysis results (Fig. 2) are checked for errors. Data set is indicated in table # 1. The default is years, air temperature, water evaporation and relative humidity. [7,15]

Table # 1

	Yillar	Havo harorati	Suv bug'lanishi	Nispiy namlik
0	1980	-5.1	14.26	80
1	1980	-4.8	20.57	72
2	1980	3.6	61.84	58
3	1980	16.9	154.84	51
4	1980	23.3	243.55	42

An analysis for the presence of noise is made in relation to the table. Noises are numbers that suddenly drop during analysis. Also, data that significantly different from other data points. For this reason, each columns is checked to ensure that it is performing correctly. (Figure 3.) All comparative results are presented below. [8]

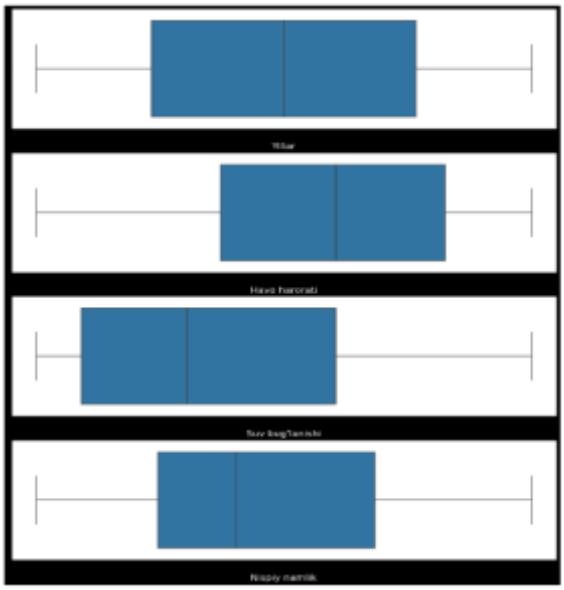

Fig. 3. Noise analysis process

After removing noises, the relative humidity and air temperature are compared with each other along the x and y axes based on the values given in the table. As a result, it can be seen that the temperature and relative humidity along the x and y axes increase from year to year. (Fig.4.5)

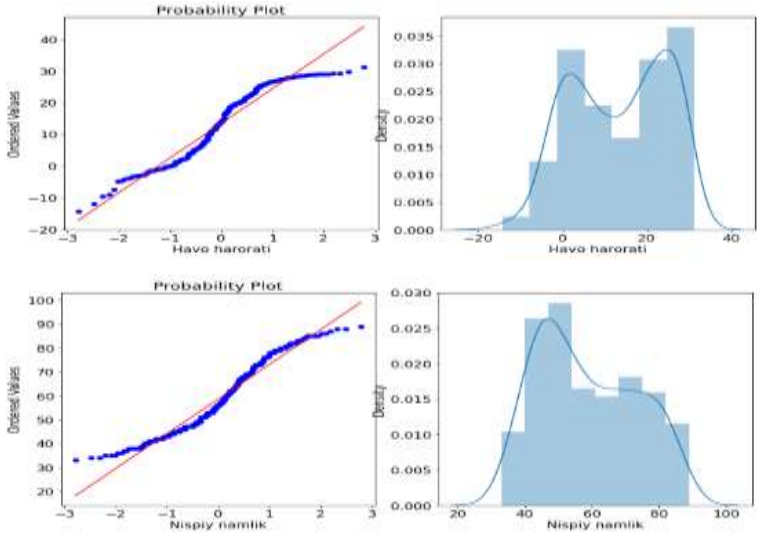

Fig.4. Graph of data distribution in columns.

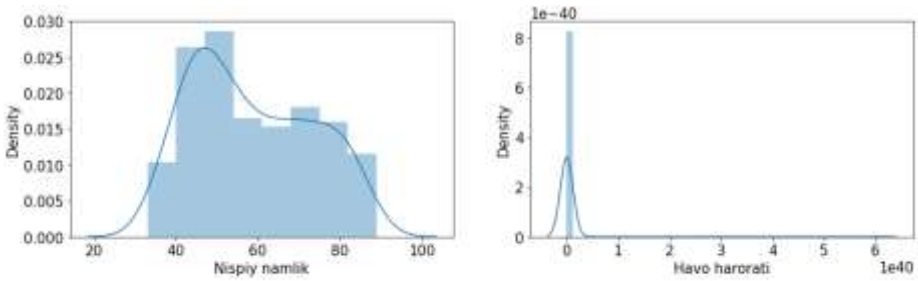

Fig.5. Graph of data distribution in columns.

As a result, you can see the result of comparision of actual and predicted results. (Fig. 6.). With the development of multiple linear regression, it is possible to predict, year after year, how many liters or tons of water will evaporate and how authorities the balance of drinking water.

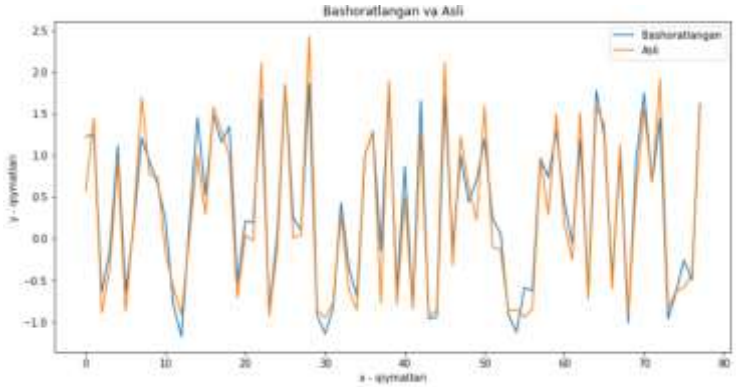

Fig. 6. Conducting a Linear Regression Outcome

DISCUSSION

Regression models describe the linear relationship between the selected variable and the dependent variables. Linear regression models use a straight line, while logistic and nonlinear regression models use a curved line. Regression predicts how an involuntary variable will change with the change of the free variables. The following is a multi-variable linear regression formula.

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$ (1)

Y - is an arbitrary variable, B0 - is a free limit, X - is an angle coefficient of X, and X - is an arbitrary variable.

In the case of linear regression, the angle coefficient determines the effect of these arbitrary variables on the arbitrary variable. It follows that the greater the angular coefficient of an arbitrary variable, the greater its degree of influence on an arbitrary variable. The objective of the linear regression model is B0, B1, B2…. Finding Bk is. We used the Ordinary least-squares (OLS) method to determine the optimal coefficients.

Linear regression models show more accurate results when the training sample is noise-free and the most relevant columns are selected. Therefore, the above linear regression model was first processed before construction.

In the multi-variable linear regression model used, groundwater level rise was selected as an involuntary variable, and air temperature, precipitation, and water evaporation were selected as independent variables.

Based on the above formula, a multidimensional linear regression model structure was constructed.

Fig.7. Structural modeling

CONCLUSION

In conclusion, it can be pointed out that without the human factor, it is possible to predict what changes will occur in nature in the future. This can be seen from the results of the above analysis. If science and technology continue to develop, natural hazards and droughts can be prevented. Thanks to the development of artificial intelligence, good results can be achieved in various fields.

Today, agriculture is one of the most important sectors of the economy. The efficiency of agricultural production is also directly related to soil and air moisture. This study aims to predict the level of water evaporation, through which we can achieve a number of successes in increasing productivity by predicting the state of soil and air moisture.

In conclusion, it is possible to predict what changes will take place in nature in the future without the human factor. This can also be seen from the results of the above analysis. Thanks to the development of artificial intelligence, good results can be achieved in various fields.

ACKNOWLEDGEMENT

This work was supported by the Tashkent University of Information Technologies and the laboratory of geo-information technologies.

REFERENCES

1. Rahmani, F., Lawson, K., Ouyang, 545 W., Appling, A., Oliver, S., and Shen, C .: Exploring the Exceptional Performance of a Deep Learning Stream Temperature Model and the Value of Streamflow Data, Environ. Res. Lett., Https://doi.org/10/ghsw9p, 2020.

2. Rajaee, T., Ebrahimi, H., and Nourani, V .: A Review of the Artificial Intelligence Methods in Groundwater Level Modeling, Journal of Hydrology, https://doi.org/10/gfvfg3, 2019.

3. Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., and Gratzki, A .: A Central European Precipitation Climatology - Part I: Generation and Validation of a High-Resolution Gridded Daily Data Set (HYRAS), Meteorol. Z., p. 22, https://doi.org/10/f5gf49, 2013.

4. Reback, J., McKinney, W., Jbrockmendel, Bossche, JVD, Augspurger, T., Cloud, P., Gfyoung, Sinhrks, Klein, A., Roeschke, M., Hawkins, S., Tratner, J ., She, C., Ayd, W., Petersen, T., Garcia, M., Schendel, J., Hayden, A., MomIsBestFriend, Jancauskas, V., Battiston, P., Seabold,

S., Chris -B1, H-Vetinari, Hoyer, S., Overmeire, W., Alimcmaster1, Dong, K., Whelan, C., and Mehyar, M .: Pandas-Dev / Pandas: Pandas 1.0.3, Zenodo, https: //doi.org/10.5281/ZENODO.3509134, 2020.
5. Région Alsace - Strasbourg: Bestandsaufnahme Der Grundwasserqualität Im Oberrheingraben / Inventaire de La Qualité Des Eaux Souterraines Dans La Vallée Du Rhin Supérieur, 1999.
6. Shen, C .: A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists, Water Resour. Res., 54, 8558-8593, https://doi.org/10/gd8cqb, 2018.
7. Sudheer, K. P., Nayak, P. C., and Ramasastri, K. S .: Improving Peak Flow Estimates in Artificial Neural Network River Flow Models, Hydrological Processes, 17, 677-686, https://doi.org/10/b39k4k, 2003.
8. Law of the Republic of Uzbekistan "On Informatization", 11.12.2003, No. 560-II, Tashkent Collected Legislation of the Republic of Uzbekistan, 2014, No. 36.
9. Resolution of the President of the Republic of Uzbekistan No. PP-4642 of 03.17.2020, "On measures for the widespread introduction of digital technologies in the city of Tashkent" National database of legislation, 18.03.2020, No. 07/20/4642/0328
10. Bezruchko B. P., Smirnov D. A. Mathematical modeling and chaotic time series. - Saratov: GosUNTS "College", 2005. - ISBN 5-94409-045-6
11. Rustam Yakhshibaev, Boburkhon Turaev, Khudoyorkhon Jamolov, Nozima Atadjanova, Elena Kim, Nargiza Sayfullaeva - International Conference on Information Science and Communications Technologies (ICISCT) 2021
12. Dzhumanov Zh.Kh., Razhabov F.F., Kim E.V., Yakhshiboev R. Development of a model and calculation of the biosignal flow balance based on microcontrollers, ISSN2181-7812 2020. P.187-188.
13. Dzhumanov Zh.Kh., Yusupov R.A., Egamberdiev Kh.S., Yakhshiboev R.E. The use of geographic information systems to substantiate promising areas for the organization of prospecting and exploration works (for example, for water supply of national economic facilities) IICS-2020.
14. Dzhumanov Zh. Kh. Yakhshiboev R. Development of a mathematical model and a software package for calculating items of the balance of information flow (for example, drinking water). IICS-2020
15. Yakhshiboev Rustam Erkinboy ўғli “Forecasting groundwater quality using machine learning” - “EDUCATION AND SCIENCE IN THE XXI CENTURY” 5-20. Page 1131.2021yy