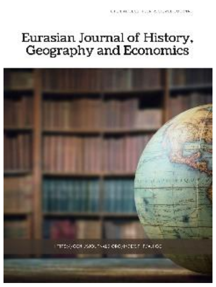


Testing the accuracy of the geographical map model using spatial statistical methods: Climatic map as a model

Dr. Ziyad Muhammad Hamid
Ministry of Education \ Kirkuk Education Directorate
dr_azhir2017@uokirkuk.edu.iq

ABSTRACT

Testing the model is an important step taken to detect the relationships and correlations between the elements of the phenomenon studied in order to measure the accuracy of the spatial models of the cartographic models. This can be implemented through the use of spatial statistical tools that have the ability to deal with geographical databases (Khair, 2002, p. 340). The geographical outputs of any geographical phenomenon depend on two things: the first is the availability of data recorded within the theoretical and practical studies and their validity and accuracy of recording. The second is the method used in the processing methods, which are chosen according to what is requested from the scholars about the type of phenomenon and the nature of its existence. It was necessary for us to choose the most appropriate methods to complete the technical work and produce the closest results in terms of validity and acceptance.

Keywords: Spatial statistical methods, simple linear regression, correlation coefficient, inevitable methods, geo-statistical, model.

1.Statistical testing:

To perform a statistical analysis to test the accuracy of the model used, Multiple Linear Regression must be relied upon, as this model is an advanced statistical method that ensures accurate inference in order to improve research results. This can be obtained through the optimal use of data in creating causal relationships between the phenomena under study (Schmidheiny, 2018, p. 2).

The scientific method of models depends on taking the data available in certain places (climatic stations), then predicting the required data in the study area where there are no measurements. These mathematical processes are called spatial statistics, which use computer mathematical models based on mathematical statistical method. This depends on the distribution of the geographical phenomenon. There follows the methodology of statistical analysis in order to predict the completion of maps of heat and rain (Al-Azzawi, 2019, p. 229).

After entering the climatic data pertaining to selected stations and according to the date of the station's opening, statistical methods are used to process those data in order to make climatic models for the elements of heat and rain. The EXCEL program is also used to extract the rates of climatic cartographic models. The spatial completion in the GIS program is one of the geostatistical methods used in preparing the required climate models. Also used here is the SPSS program in order to obtain the spatial correlation of the data (Al-Farra, 1987, p. 215).

2. Simple linear regression:

It is a statistical method that enables us to build a statistical model to estimate the relationship between one quantitative variable, which is the dependent variable and an independent quantitative variable. This is called simple regression while the model with several independent quantitative variables is called multiple regression (Al-Atbi & Al-Taei, 2013, p. 206). Multiple linear regression is not just a single method, but rather a set of methods that can be used to find out the relationship between a continuous dependent variable and a number of independent variables that are usually continuous. The linear equation in multiple linear regression runs follows (Tranmer & Elliott, 2001, p. 21):

$Y = a + b1X1 + b2X2 + \dots\dots\dots + e$

Y = dependent variable

A = Constant or Intercept

B1 = slope of the regression y on the first independent variable

B2 = slope of the regression y on the second independent variable

1X = the first independent variable

2X = the second independent variable

After obtaining the results of the regression equation, these results must be subjected to a set of statistical tests in order to verify that these coefficients are statistically acceptable, i.e. statistically significant (Al-Jalouni, 2008, p. 241). To test the validity of the model and its efficiency in interpreting the relationship between the independent variables and the dependent variable, a set of statistical tests will be adopted, which we will discuss in some detail. Our study is limited to studying the relationship between two variables, namely, the methods of spatial completion and the strength of the correlation with the total area of each method.

3. The multiple correlation coefficient (R):

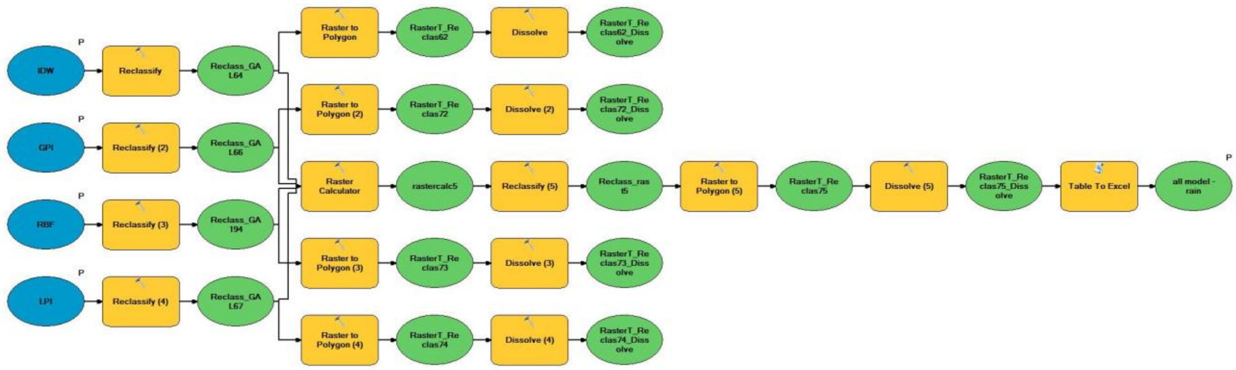
It is the multiple correlation coefficient, which measures the strength of the relationship between two or more variables. The sign of the correlation coefficient here does not indicate the direction of the relationship because this direction is not uniform for all variables, and is determined by the regression coefficients within the linear regression model (Al-Atbi & Al-Hiti, 2011, pp. 172-173).

4. Adjusted R-squared R2 coefficient of determination:

The squared linear correlation coefficient, symbolized by R2, and its value ranges between one and zero (Al-Atbi & Al-Taei, 2013, p. 220), This test is used to measure the strength of the estimated model by determining the percentage of changes in the dependent variable (Y) the total area shown by the regression model, and its value falls between one and zero. The interpretive strength of the model increases as its value increases and vice versa (Shehadeh, 2002, p. 383).

This statistical model was used as it expresses the degree of influence of the indicators used for each method through the areas that were extracted after finishing the spatial completion process for each method, and determining which factors have the most influence on it. All models were built using deterministic tools, based on Spatial Interpolation and based on giving a new layer in the form of Raster resulting from the layer of climatic stations and inserting the layer and then classifying the input layer through Reclassify, into five classes. This process is applied to all layers and then these layers are merged using Map algebra within the raster calculator in spatial analyst, GIS. This allows two or more network layers (maps) of similar dimensions to produce a new mesh layer (map) using Algebraic processes such as addition, subtraction, etc.(Longley et al., 2005, pp 414-417) and then building the Model tool in the ARCGIS environment, using the Model builder. There follows the Algebra-based Raster tool (Map Algebra), in addition to the (Wight overly) tool. As in Figure (1) and (2).

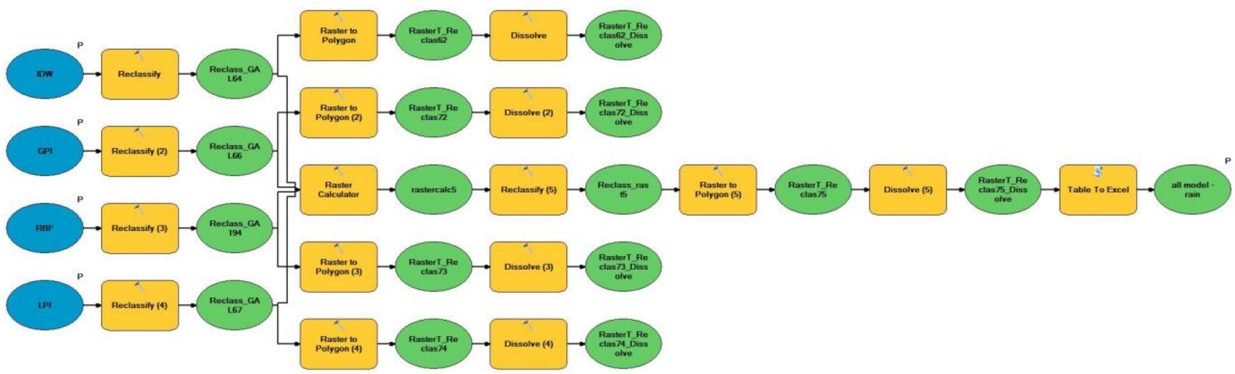
Figure (1): A model for extracting the areas of the inevitable roads for rain.



Source: Done by the researcher based on the environment of the ARC GIS 10.6.1 program.

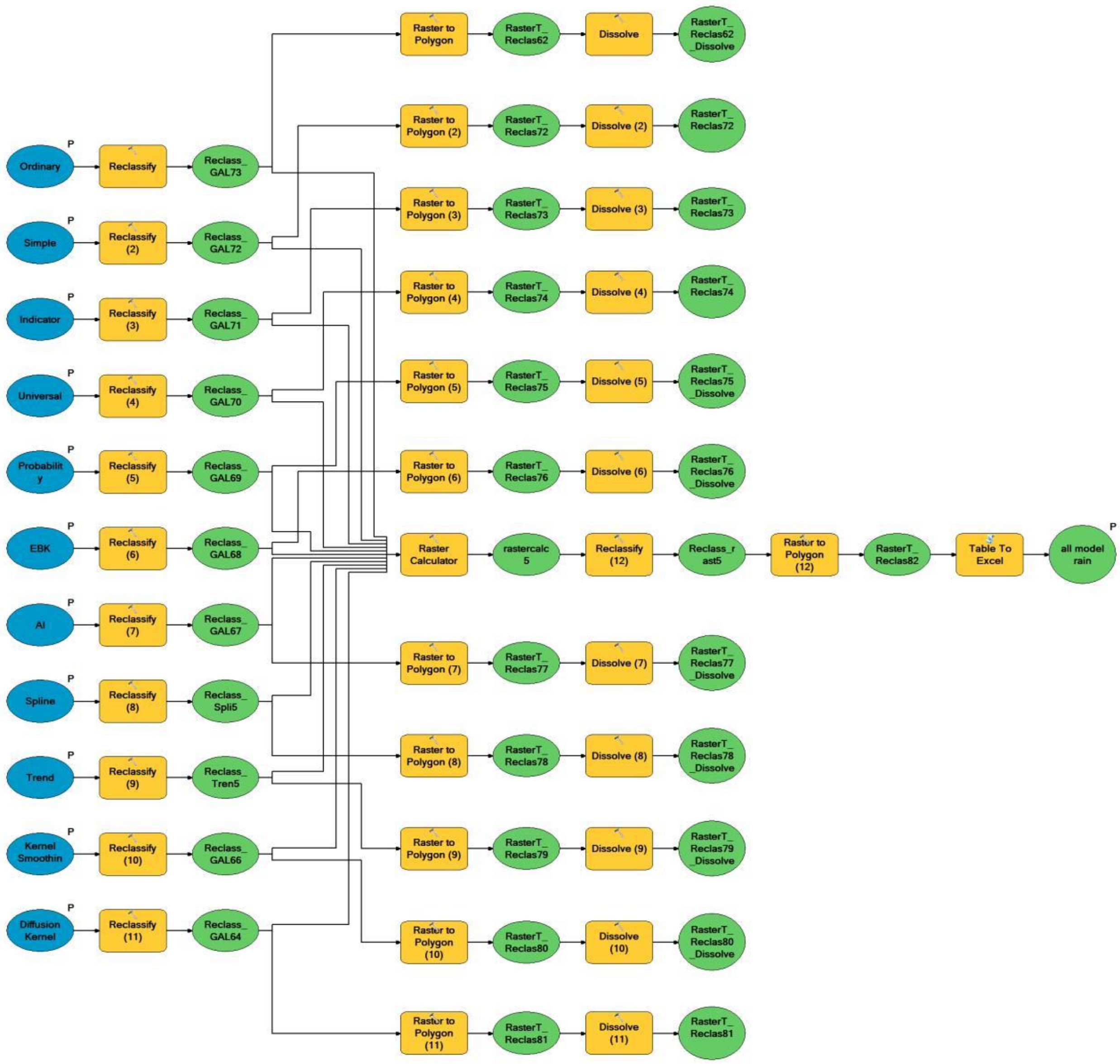
Figure (2): A model for extracting the areas of the inevitable roads for rain.

Source: Done by the researcher based on the environment of the ARC GIS 10.6.1 program.



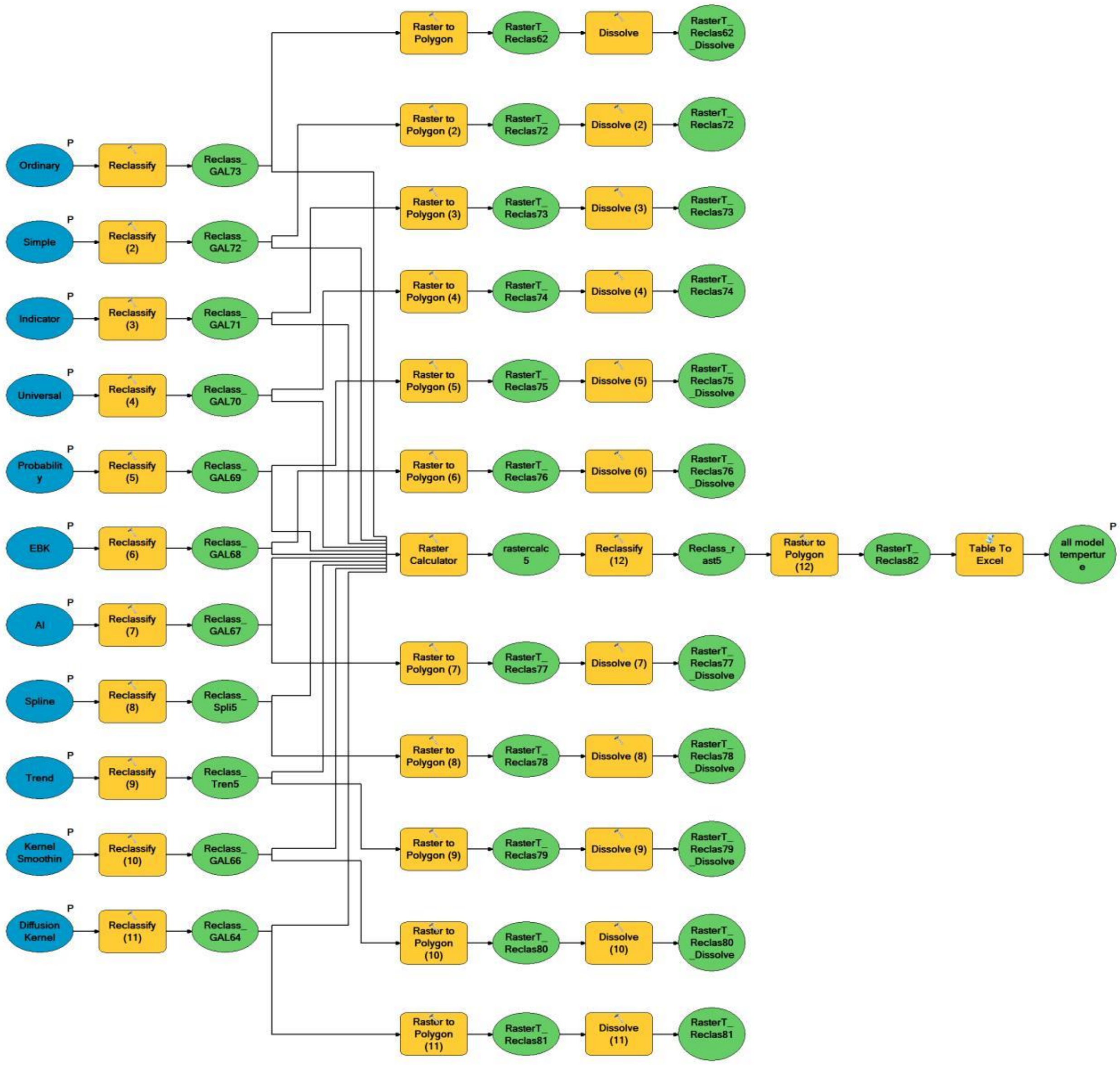
The same steps were followed in the methods of geostatistical tools, which are based on spatial interpolation. They are also based on giving a new layer in the form of Raster resulting from the layer of climatic stations and inserting the layer, and then classifying the entered layer through (Reclassify) into five categories. This process is applied to all layers. These layers are then merged using Map algebra within the raster calculator in spatial analyst. Then, we build the Model tool in the ARCGIS environment using Model builder and then the (Raster) tool based on cartographic algebra (Map Algebra). These are in addition to the (Wight overly) tool as can be shown in Figure (3) and Figure (4).

Figure (3): A model for extracting areas, geo-statistical methods for temperature.



Source: Done by the researcher based on the environment of the ARC GIS 10.6.1 program.

Figure (4) Area extraction model for geostatistical methods of rain.



Source: Done by the researcher based on the environment of the ARC GIS 10.6.1 program.

The results of the model were applied based on the SPSS program, and the results of the application are shown below. As in Table (1) and (2) and as follows:

We conclude from the above the reliability of the estimated model through the values of statistical tests. It refers to the value of the multiple correlation (R) and the strong relationship between the parameters and the dependent variable, as it reached (0.59) for the deterministic method (RPF). This is the highest correlation strength for the rate of heat, while the lowest correlation strength recorded was for the method (GPI) with a correlation strength of (0.44). As for the value of the interpretation coefficient (R2), which reached (0.080), the highest correlation strength recorded was for (LPI) and the lowest correlation strength recorded at a rate of (0.22) (GPI) from the total area of the collected map.

The model estimated through the values of statistical tests shows that rain is reliable, as the multiple correlation (R) value of (0.88) for the (GPI) method indicates that the relationship is strong between the determination area and the total area. The least correlation strength was (0.49) for the (LPI) method while the value of the interpretation coefficient (R2), which reached (0.78), had the highest correlation strength for the (GPI) method. The lowest correlation strength (0.24) was the (LPI) method.

Table (3) and (4) show the reliability of the estimated model through the values of statistical tests, as the value of the multiple correlation (R) of the geostatistical surface indicates that the relationship is strong between the parameters and the dependent variable. It reached (0.96) for the statistical surface (DK), which is the highest correlational strength of the rate of heat, while the lowest was recorded for (ORH) method, with a correlation strength of (0.18). As for the value of the interpretation coefficient (R2), which is (0.92), the highest correlation strength recorded was for (DK), and the lowest recorded at a rate of (0.09) was for (EBK) from the total area of the collected map of heat.

As for the correlation strength of the geostatistical surface of rain, it shows the reliability of the estimated model through the values of statistical tests. The value of the multiple correlation (R) of (0.99) for the (Spline-ok) method indicates that the relationship is strong between the completion area and the total area. The lowest correlation strength reached (0.57) for the (Dk) method, and the value of the interpretation coefficient (R2) which reached (0.98) had the highest correlation strength for the (Spline-OK) method, while the lowest correlation strength (0.33) was the (DK) method

.

Table (1): the results of applying the linear regression equation to the deterministic methods of the rate of heat

	Methods	Combined space	(Constant)	X	(R)	(R2)	Sig
Estimated equation	IDW &	Y=	23,460	0.469	0.52	0.28	0.050
	GPI &	Y=	18252	0.143	0.44	0.22	0.037
	& RPF	Y=	26505	0.659	0.59	0.35	0.023
	LPI &	Y=	18550	0. 401	0.49	0.26	0.080

Source: Based on program outputs Spss.

Table (2): Results of the application of the linear regression equation to deterministic methods for the rate of rain

	Methods	Combined space	(Constant)	X	(R)	(R2)	Sig
Estimated equation	IDWR	Y=	29507.3	2.2	0.83	0.68	0.050
	GPIR	Y=	22053.9	2.6	0.88	0.78	0.045
	RPFR	Y=	35157.9	1.8	0.87	0.75	0.55
	LPIR	Y=	67401.4	0.219	0.49	0.24	0.080

Source: Based on program outputs Spss.

Table (3): The results of applying the linear regression equation of geostatistical methods to the mean temperature

	Methods	Combined space	(Constant)	X	(R)	(R2)	Sig
Estimated equation	PRH	Y=	108441.5	4.2	0.72	0.52	0.042
	UNH	Y=	227955.4	3.2	0.47	0.22	0.023
	INH	Y=	152972.3	1.4	0.24	0.12	0.068
	SMPH	Y=	127945.3	2.9	0.35	0.17	0.058
	ORH	Y=	162404.4	0.8	0.18	0.3	0.76
	DKH	Y=	103665.7	4.5	0.96	0.92	0.008
	KSH	Y=	79244.7	6.03	0.87	0.75	0.050
	TH	Y=	123452.3	3.2	0.70	0.49	0.030
	SH	Y=	162094	0.85	0.32	0.10	0.590
	AIH	Y=	150714.6	1.56	0.43	0.18	0.469
	EBKH	Y=	157855.4	1.11	0.31	0.09	0.611

Source: Based on program outputs Spss.

Table (4): The results of applying the linear regression equation to the geostatistical surface of the rainfall rate

	Methods	**Combined space**	**(Constant)**	**X**	**(R)**	**(R2)**	**Sig**
Estimated equation	EBK	**Y=**	**-359.7**	**0.71**	**0.97**	**0.96**	**0.000**
	AL	**Y=**	**693.3**	**0.93**	**0.94**	**0.88**	**0.049**
	KS	**Y=**	**-10016.0**	**1.59**	**0.80**	**0.64**	**0.195**
	DK	**Y=**	**36619.3**	**-0.95**	**0.57**	**0.33**	**0.424**
	Trend	**Y=**	**39095.4**	**-1.14**	**0.92**	**0.84**	**0.029**
	Spline	**Y=**	**4431.1**	**0.697**	**0.99**	**0.98**	**0.005**
	Ordinary	**Y=**	**235.5**	**0.99**	**0.99**	**0.98**	**0.004**
	Simple	**Y=**	**-8271.7**	**1.49**	**0.73**	**0.54**	**0.265**
	Indicator	**Y=**	**35370.7**	**-0.94**	**0.98**	**0.96**	**0.004**
	Universal	**Y=**	**-74342.0**	**5.92**	**0.86**	**0.75**	**0.132**
	Probability	**Y=**	**608.2**	**1.57**	**0.98**	**0.97**	**0.010**

6. Conclusion

The geographical map is subject to multiple verifications and tests, the aim of which is to ensure the validity of the results and their conformity with the actual situation. These processes are conducted using many methods, technical methods and statistical methods, each according to the nature of the phenomenon to be examined, and the geographical aspect that was studied here. In this study, which was about proving the accuracy of the mapping models of climate elements, spatial statistical methods were used to calculate the areas by deterministic and geo-statistical methods. This is where the algorithmic equations were applied using GIS techniques. Also applied was the linear regression equation in both ways where the results were close to acceptance level and with high perceptive strength to match the phenomenon studied and according to the data recorded from the ground stations available in the region for the elements of heat and rain.

References

1. Al-Atbi, S. A., & Al-Hiti, M. Y. (2011). *Scientific Research Methodology*. Baghdad: Asdiqaa Press.
2. Al-Atbi, S. A., & Al-Taei, I. A. (2013). *Statistics and Geographical Modeling*. Baghdad: Akram Printing Press.
3. Al-Azzawi, A. A. (2019). Comparison of spatial interview techniques for groundwater staples in Tal Afar, GIS.. *Journal of Kirkuk University, 14*(1), 226-244.
4. Al-Farra, M. A. (1987). *Research Methods in Geography by Quantitative Means* (3rd ed.). Kuwait: Publications Agency.
5. Al-Jalouni, F. K. (2008). Multiple Linear Regression Analysis Method in Studying the Most Important Economic, Social and Demographic Variables Affecting the Total Birth Rate.. *Damascus University Journal of Economic and Legal Sciences, 24*(2).
6. Khair, S. (2002). *Geography: Its subjects, methods, and objectives*. Beirut: Contemporary Thought House.
7. Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). *Geographic information systems and science*. John Wiley & Sons.
8. Schmidheiny, K. (2018). *The Multiple Linear Regression Mode, Short Guides to Micro econometrics*. Basel: Universitaet Basel.
9. Shehadeh, N. (2002). *Quantitative Methods in Geography using Computer* (2nd ed.). Amman: Dar Al-Safa Publishing and Distribution.
10. Tranmer, M., & Elliot, M. (2001). Multiple linear regression. *The Cathie Marsh Centre for Census and Survey Research (CCSR)*, *5*(5), 1-5.