Technium Vol. 4, No.10, pp.115-125 (2022)
ISSN: 2668-778X
www.techniumscience.com

Comparative Evaluation of Prediction Model between Inference Fuzzy System and Universal Kriging for Spatial Data

Ghanim Mahmood Dhaher [1*], Ibrahim Abdulghany Ibrahim [2]

[1] Department of Mathematics, University of Mosul, Iraq College of Education for Pure Sciences
[2] Department of Computer Science, University of Mosul, Iraq College of Education for Pure Sciences
*Corresponding author email: **ghanim-hassod@uomosul.edu.iq**

Abstract. This paper dealt with one of the spatial interpolation methods in the geostatistics field. The purpose of this research is to get the parameters of unbiased estimators based on regionalized random variables in spatial statistics. In this paper, we used universal kriging with the fuzzy inference system by the Mamdani technique. the objective of this work is to estimate the parameters of covariance functions relying on spatial real for the depth of groundwater in Mosul city, Iraq. The data adopted contains (100) real data with locations representing the depth. From the results we show the best model with the constructs of weights, we illustrate the performance of universal kriging is the best when corresponding with the fuzzy system. In conclusion, the improvement of any method of spatial interpolation or fuzzy system does not depend on more statistical structures but depends on the efficiency of the method which satisfies the conditions of weights and minimum variance errors. All programming is applied by Matlab language.

Keywords. spatial statistics, fuzzy inference system, universal kriging, depth of groundwater.

1. Introduction

The concept of fuzzy set theory was started by Lotfi A. Zadeh in 1965 when he published his famous research on "fuzzy set". L. A. Zadeh linked the probability theory with the fuzzy sets to lead the mathematical logic. After Zadeh, many studies introduce some concepts of fuzzy set theory dealt with the comparison between the fuzzy inference systems as Mamdani and Sugeno [1], while the other studies were interested in prediction using fuzzy systems [2], [3]. Geostatistics is a branch of statistics sciences, that is interesting to study the regionalized variables which are related the spatial prediction. In (1951) in mining engineering, Krige create the method of prediction process from gold concentrations in the ore of South Africa. Mining engineer D. G. Krige was the first published in spatial statistics, later, a French mathematician G. Matheron developed the method of kriging based on the master thesis of D. G. Krige in geostatistics. Spatial interpolation can be defined as one of the statistical methods for the prediction process for spatial data, where many researchers and statisticians use graphics, curves, and charts that help us give an idea of spatial variation or spatial distribution in the prediction process. There are many different types of interpolation methods, including linear interpolation and multivariate interpolation. Many studies use prediction with variogram functions such as [4], [5], [6]. Matheron was the first scientist to introduce the prediction of spatial statistics using the parameters of covariance functions.

2. Methods and Materials 2.1 Fuzzy Theory

Lotfi A. Zadeh was the first scientist to define the fuzzy set. Fuzzy set A is defined as: $A = \{x, \mu_A(x), x \in X\}$, where $\mu_A(x)$ is the membership degree of x and X is collection of element numbers x, and $\mu_A: X \rightarrow [0,1]$. Fuzzy logic is the theory of fuzzy sets, where the fuzzy logic relies on the idea of degrees of things such as (height, speed, …, etc.) [7].

2.2 Mamdani Fuzzy Logic (MFL)

The scientist Ibrahim Mamdani of London university give the first fuzzy system and known as the Mamdani style. The output of the membership function in Mamdani is fuzzy sets, therefore the output variables need processes through the inputs to prediction. There are four steps of Mamdani fuzzy logic defined as the following: [2[, [8].

Step 1: Fuzzification. The purpose of fuzzification is to map the inputs. The crisp input is a numerical value, fuzzy set [Lower, Median, and High].

Step 2: Rule Evaluation. Inputs are applied to a set of (If/Then) control rules The fuzzy operator (AND or OR) is used to get a single number and to obtain the membership function.

Step 3: Aggregation of the rule output

Is the process of all outputs of all rules in step 2, for each output variable the result is one fuzzy set.

Step 4: Defuzzification

The final step of (MFL) is defuzzification where the output has to be a crisp number. There are several defuzzification methods such as (the centroid technique, center of gravity, …, etc.)

$$COG = \frac{\int_a^b \mu_A(x)x dx}{\int_a^b \mu_A(x) dx} \quad (1)$$

The hypothetical fuzzy variogram uses the parameters of variogram function, such as nugget effect, partial variance, and range defined in a fuzzy spherical variogram and also defined as a triangular membership function:

$$T(x/a,b,c) = \left\{ \begin{matrix} 0 & x \leq a \; or \; x \geq c \\ 1 & x = b \\ \frac{x-a}{b-a} & a < x < b \\ 1 - \frac{x-b}{c-b} & b < x < c \end{matrix} \right\} \quad (2)$$

Where $a \leq b \leq c$, [3], [9].

2.3 Kriging Technique

Ordinary kriging is the most important prediction in spatial statistics. The prediction based on the model:

$$Z(x) = \mu + \epsilon(x) \quad (3)$$

$$\hat{Z}(x_o) = \sum_{i=1}^{n} \alpha_i(x_o).Z(x_i) = \alpha_0^T Z \quad (4)$$

Where α_i is i weights, and Z is the vector of n observation with location.

The predictor $\hat{Z}(x_o)$ of random variable $Z(x)$ is said to be unbiased if

$$E[Z(x) - \hat{Z}(x_o)] = 0$$

And the unbiasedness for any choice of α and the constant ω_o can be choose as

$$\omega_o = E[Z(x_o) - \alpha^T E(Z(x))] = \mu_o - \alpha^T \mu$$

$= \mu_o - \sum_{i=1}^{n} \alpha_i \mu_i$, where μ_o is expected value of $Z(x_o)$ and $\mu_o = E[Z(x_I)]$, [10], [11].

2.4 Universal Kriging

Spatial data can be through resulting from regionalized random variables in the study field on stochastic process: $Z = \{Z(x): x \in D\}$

Universal kriging assume the model $Z(x) = \mu(x) + \in(x),\ x \in D \subset R^2$

Where $\mu(x)$ is some deterministic function and $\in(x)$ is random variation is the errors with the mean is zero. Assume $Z(x_i), i = 1,2,\dots,n$ to be regionalized variables at the points (x_i), are locations at random field.

Let $Z(x) = \sum_{i=1}^{n} \beta_{i-1} f_{i-1}(x) + \in(x)$ (5)

Where $\in(x)$ is the zero mean, f_{i-1} is a basic functions be known, and β_{i-1} is coefficients to be estimated. And the universal kriging estimator at (x_o), defined as:

$$\hat{Z}_{UK}(x_o) = \sum_{i=1}^{n} \alpha_i . Z(x_i) = \alpha_{UK}^{T} Z = \alpha_{UK}^{T} \gamma + f_o^{T} \alpha \quad (6)$$

And the variance of universal kriging denoted as σ_{UK}^2 and calculate as :

$$\sigma_{UK}^2 = var\left(\hat{Z}(x_o) - Z(x_o)\right) = -\alpha_{UK}^{T} \tau\, \alpha_{UK} + 2\alpha_{UK}^{T} \gamma$$

we can obtain the minimum universal kriging as the following:

$$\sigma_{UK}^2 = \sum_{i=1}^{n} \alpha_{UK}^{T} \gamma z(x_i - x_o) + \sum_{l=1}^{L} \beta_{l-1} f_{l-1}(x_o) \quad (7)$$

[12], [13].

2.5 Variogram Function

The random variables are called second-order stationarity with mean μ and covariance C(h) if:

$i)$ $\mu \in R$ of $Z(x)$ is constant, i.e. $E[Z(x)] = \mu(x) = \mu,\ \forall x \in D$

$ii)$ $C(h) = cov\big(Z(x), Z(x+h)\big) = E[Z(x), Z(x+h) - \mu^2$

When the $Z(x)$ is satisfy the second-order stationarity then:

$$\gamma(h) = C(0) - C(h)$$

Where $C(0)$ is the variance at (0), [4].

The theoretical variogram function defined by Matheron and Cressie denoted as $2\gamma(h)$ where:

$$2\gamma(h) = var[Z(x+h) - Z(x)] = E[Z(x+h) - Z(x)]^2 \text{ where } x, x+h \in D$$

The variogram function assume the stationarity conditions and it's a finite doesn't rely on the location of x in domain D.

Matheron defined the method of moment estimator as:

$$\gamma(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \quad (8)$$

Where $N(h)$ is the number of all pairs of points with lag h.

To fitting a variogram function, we must getting the three most common parameters defined as the following:

- Nugget (c_o)or discontinuous at origin point,
- Sill $(c_o + c)$ or variance of variogram.
- Range is a distance on x-axis at variogram is stable

These parameters used in order to obtain the best fitting covariance model, [14], [15].

3. Application of groundwater data

3.1 Study Data

This research depends on real spatial data on the depth of groundwater in Nineveh governorate, Iraq. These data contain (100) real data with real locations. We will introduce some depth data for groundwater in Nineveh Governorate. Table (1) below shows the data for groundwater where Z is the depth and x, y are the location.

Table (1) some real data for groundwater where Z is the depth and x, and y is the location.

Z	X	Y
130	36.14493333	43.68583333
150	35.97408333	43.63805556
67	36.09078333	43.384325
150	35.9282	43.67255
⋮	⋮	⋮

3.2 Results of Kriging Techniques

Table (2): results of variogram function for depth of groundwater

G1 (0°)	G2 (90°)	G3 (45°)	G4 (135°)	G5 (0°,90°)	G6 (45°,135°)
0.1342	0.0655	0.0771	0.0550	0.0334	0.0661
0. 4137	0.2360	0.2828	0.1941	0.1201	0.2385
0.8097	0.5078	0.6104	0.4138	0.2580	0.5121
1.3753	0.8792	1.0552	0.7108	0.4465	0.8830
2.10333	1.4078	1.7097	1.1341	0.7144	1.4219
2.94582	2.0239	2.4432	1.6498	1.0267	2.0465
3.72886	2.6592	3.2650	2.1521	1.3482	2.7085
4.56364	3.5372	4.3198	2.8322	1.7914	3.5760
5.62273	4.5143	5.5284	3.5375	2.2852	4.5329
6.14364	5.6094	7.1289	4.2849	2.8354	5.7069

Table (2) describes the results of the variogram function for all theta of the compass, where G1 denoted the results of theta (0°), G2 is the results of theta (90°), G3 is the results of theta (45°), and G4 from theta (135°), while the average of variogram function by G5, and G6 of theta (0°,90°) and (45°, 135°) respectively, where all the variogram multiply (10^4)

Table (3): results of properties of the variogram function for depth of groundwater

Theta / Stat.	θ = 0°	θ = 90°	θ = 45°	θ = 135°	θ = 0°, 90°	θ = 45°, 135°
Min	0.1342	0.0655	0.0771	0.0550	0.0334	0.0661
Max	0.614	5.609e+004	7.128e+004	4.285e+004	2.835e+004	5.706e+004
Median	2.252	1.716e+004	2.076e+004	1.392e+004	8.705	1.734e+004
Range	9	9	12.73	12.73	9	12.73

Table (3) illustrates the properties of the variogram function for depth groundwater data (min, max, median, and range) with all theta of the compass, and the average of two thetas that have the same lag (or distance) between the values of depth variables.

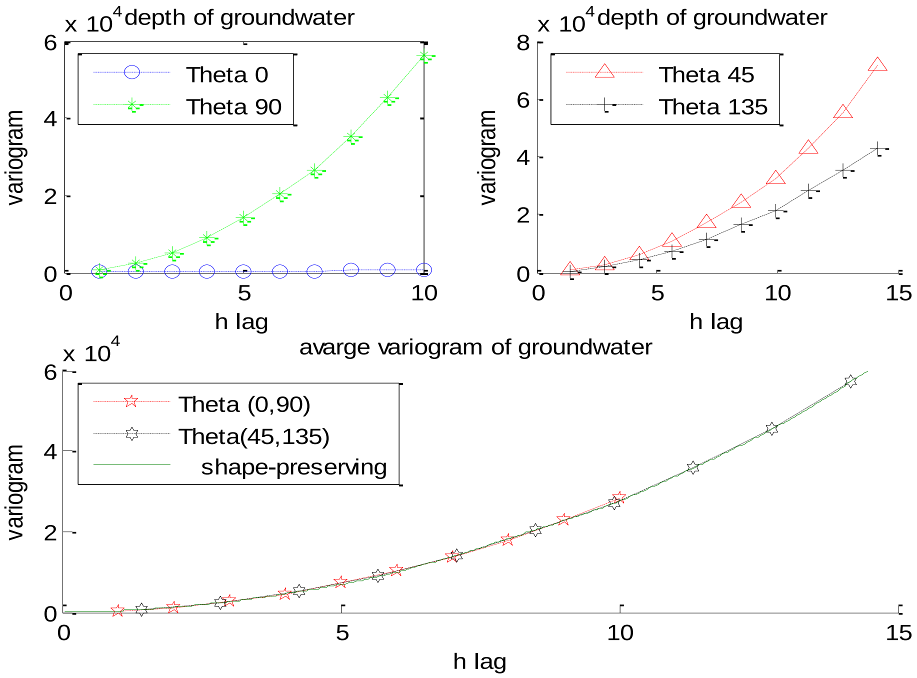

Figure (1): plots of curves variogram function for depth data

Figure (1) show the curves of the variogram function for all theta of the compass, in two up figures For theta (θ=0°, and 90°) and theta (45, 135) with the average of the variogram function (θ=0°,90°) is red curve, And the black curve of (θ=45°,135°). With the shape preserving interpolation, as indicated by three curves of average variogram function.

Table (4): results of the variogram function for logarithm depth data of groundwater

G1	G2	G3	G4	G5	G6
θ = 0°	θ = 90°	θ = 45°	θ = 135°	0°,90°	45°,135°
0.0007	0.0293	0.0343	0.0239	0.0150	0.0291
0.0028	0.0827	0.1004	0.0634	0.0427	0.0819
0.006	0.1550	0.1918	0.1154	0.0805	0.1536
0.0104	0.2494	0.3126	0.1821	0.1299	0.2474
0.0159	0.3862	0.5009	0.2794	0.2011	0.3902
0.0220	0.5530	0.7269	0.4016	0.2875	0.5643
0.0284	0.7286	0.9966	0.5216	0.3785	0.7591
0.0355	0.984	1.3763	0.6841	0.5100	1.0302
0.0429	1.3151	1.9087	0.8603	0.6790	1.3845
0.0447	1.8152	2.8296	1.0835	0.9300	1.9566

Table (4) illustrates the results of variogram function for depth groundwater data after taking the logarithm of depth, the results of G1, G2,G3, and G4) are all thetas and the average of two thetas that have the same lag (or distance) between the values of depth variables.

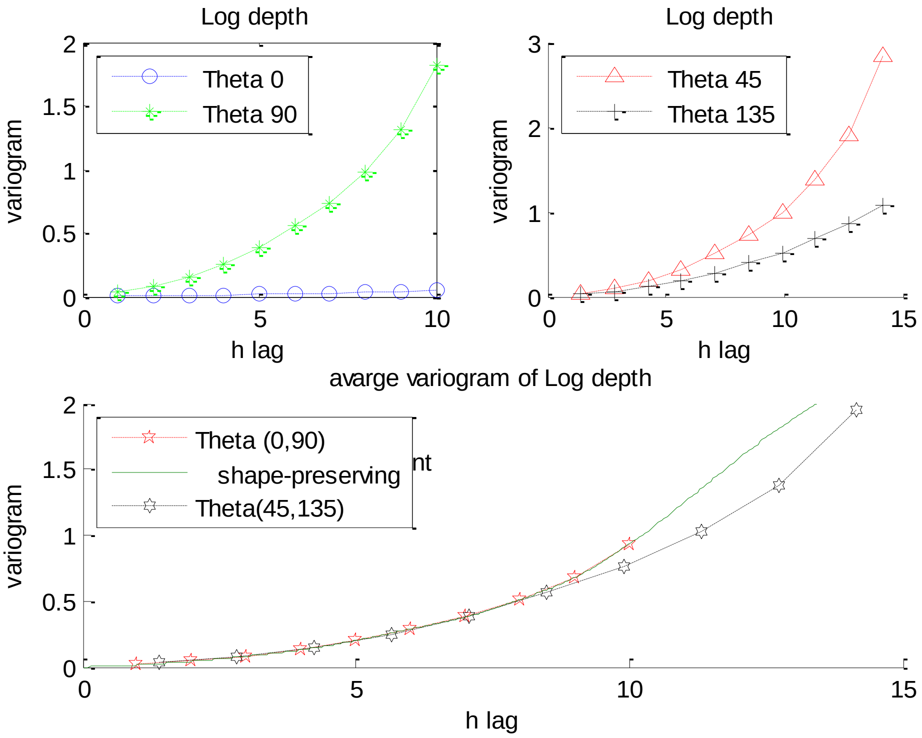

Figure (2): results of variogram function for the logarithm of depth data

Figure (2) describe the curves of the variogram function after taking the logarithms of depth data. The first figure on the up left illustrates the curves in two thetas (0°) and (90°), while the second figure is up the right of two thetas (45°), and (135°). And the average of the variogram function is in the down figure by the red curve, and black curve, while the shape-preserving by black line is the figure of the average function.

Table (5): properties of the variogram function for logarithm depth data

Theta / Stat.	θ = 0°	θ = 90°	θ = 45°	θ = 135°	θ = 0°, 90°	θ = 45°, 135°
Min	0.00074 81	0.02927	0.03432	0.02389	0.01501	0.0291
Max	0.04474	1.815	2.83	1.084	0.93	1.957
Median	0.01895	0.4696	0.6139	0.3405	0.2443	0.4772
Range	9	9	12.73	12.73	9	12.73

Table (5) show the results and properties of the variogram function after taking the logarithm of groundwater data, these properties represent min, max, median, and range for all basic theta of the compass. These properties gave good parameters for the prediction process. Through the behaviour of curves correspondence between the results of variogram function in all directions, we show that curves nearest to the power model that it's defined as the following:

$$\gamma(h) = \begin{cases} c_o + c\,|h|^w, & 0 < w < 2 \\ 0 & h = 0 \end{cases} \tag{9}$$

When the nugget effect (c_o) is clear in Table (4) the first row is in all directions of theta, and sill in the second row in the same Table(4). And the prediction of some points relies on the predictor and variance of universal kriging based on equations (6), (7). If we select the eight real data of groundwater (61, 66, 73, 84, 112, 214, 300, 326) at locations, then the work prediction in these depth data with the results of prediction (59.5, 62.7, 74.5, 84.7, 109.5, 200.8, 308.9, 330.8) we note the errors of prediction are very small and variance of universal kriging are: (0.033, 0.145, 0.205, 0.045,0.078, 0.189, 0.299,0.32).

3.3 Results of Fuzzy System

After we arranged the real data by increasing each value (X, Y, and Z), we classified it into three classes L, M, and H. To find the center of each class for inputs X, Y, and output Z (see Table (6) below).

Table (6): Centers of variables for x, y and z

	Input X		Input Y		Output Z	
Max	519.7		836		2.4150	
Min	509.6		822.7		1.7853	
Member	3		3		3	
Center 1	512.125	L	862.025	L	1.9427	L
Center 2	514.65	M	829.35	M	2.1002	M
Center 3	517.175	H	832.675	H	2.2576	H

By Matlab language, we applied the real data of levels of groundwater in fuzzy logic, we got the following:

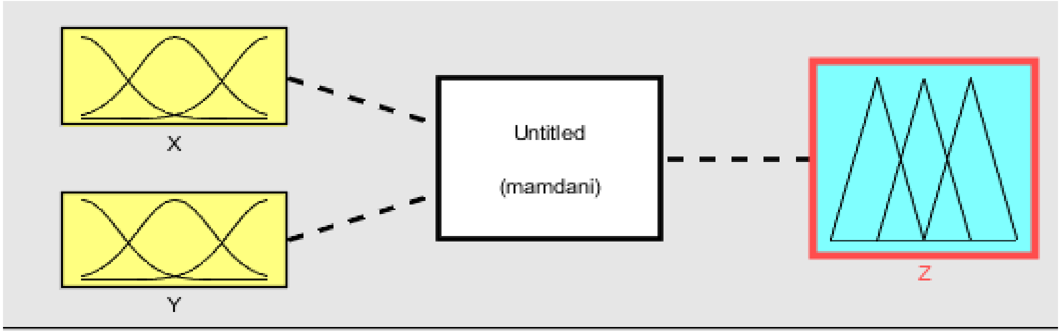

Figure (3): inputs and output of mamdani style

Figure (3) show there are two inputs X, and Y with one output Z by mamdani inference system.

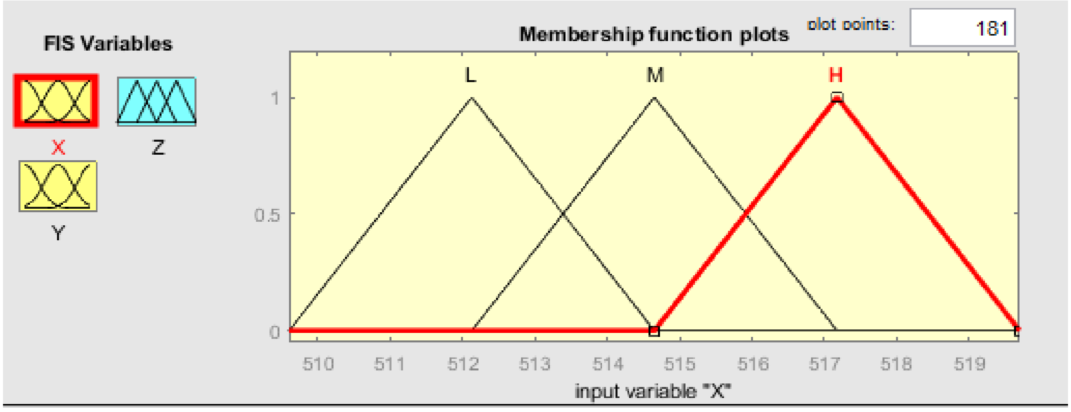

Figure (4): input variable X with membership function

Figure (4) show the input variable X with the range [509.6 519.7] and [514.6 517.2 519.7] where (517.2) is the center denoted as H (red line), with the membership function.

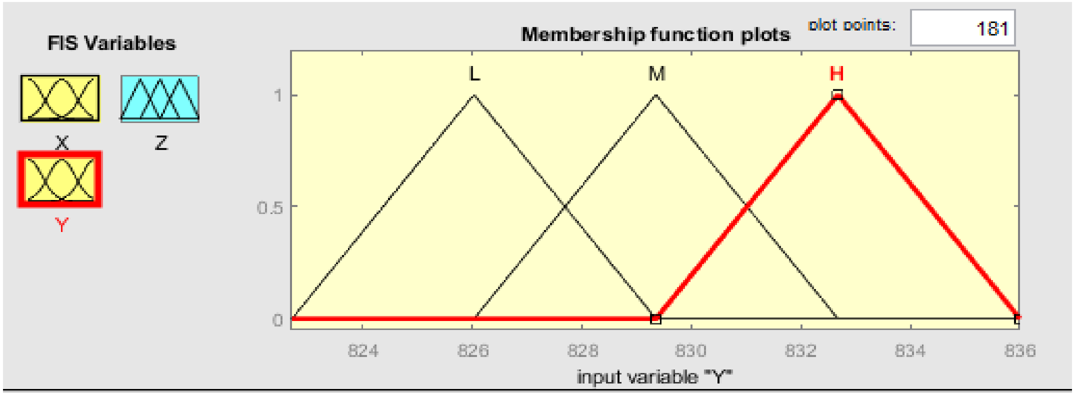

Figure (5): input variable Y with membership function

Figure (5) show the input of Y variable in the range [822.7 836] and [829.4 832.7 836] Where the center is (832.7) denoted as H (red line).

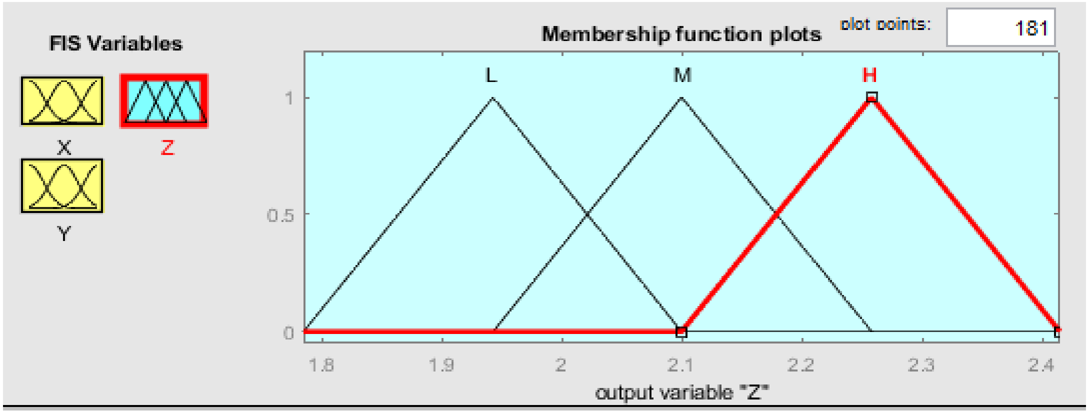

Figure (6): output variable Z with membership function

Figure (6) illustrate the output variable Z in the range [1.785 2.415] and [2.1 2.258 2.415] Where (2.258) is the center of Z denoted H (red line).
And the rules of second step was taken as the following:

1- if (X is L) and (Y is H) then (Z is H)
2- if (X is L) and (Y is H) then (Z is M)
3- if (X is L) and (Y is M) then (Z is H)
4- if (X is M) and (Y is H) then (Z is H)
5- if (X is M) and (Y is L) then (Z is M)
6- if (X is M) and (Y is M) then (Z is H)
7- if (X is H) and (Y is L) then (Z is M)
8- if (X is H) and (Y is L) then (Z is H)
9- if (X is H) and (Y is M) then (Z is M)

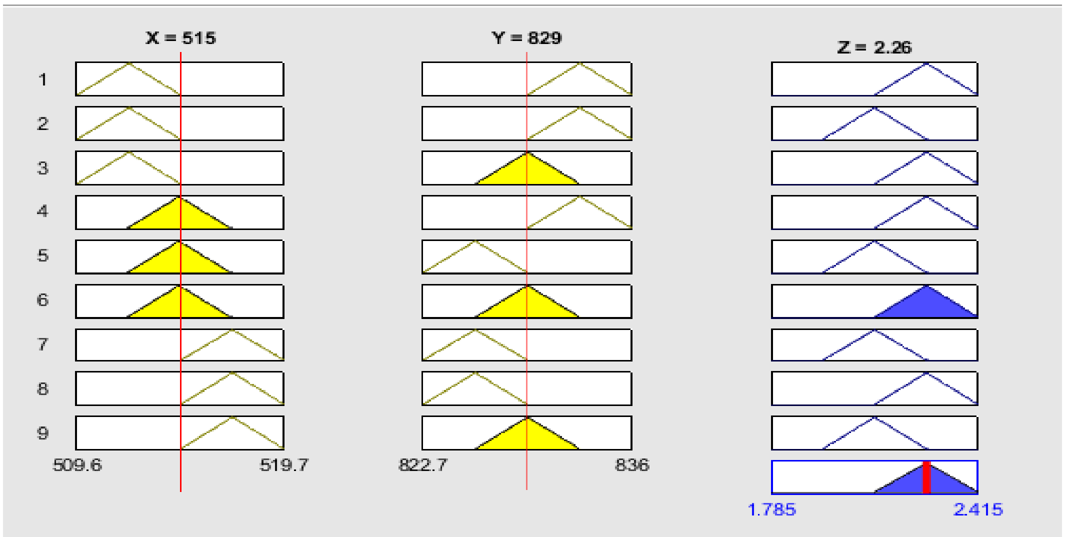

Figure (7): Output Z by using min-max

Figure (7) shows the results of output Z from the value X=515, and Y=829 we get the output of Z=2.26 by using the function min-max.

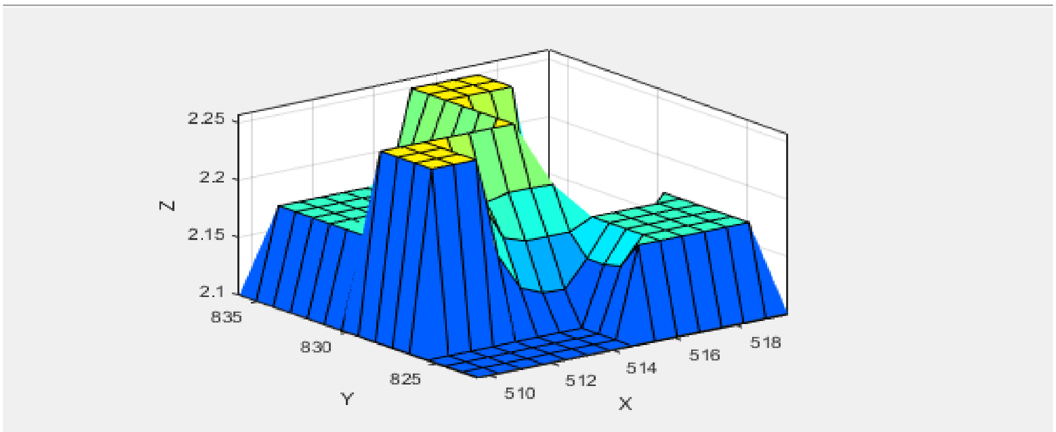

Figure (8): results of three dimensions

Figure (8) illustrate the inputs X on x-axis, and inputs Y on y-axis and values of Z on N-axis, by three dimensions.

In Fuzzification step: For example, if (X_1 = 512.2 and X_2= 514.6 , y_1 =1, y_2=0) and by using: $Y = \frac{(x-xi)(yi+1-yi)}{(xi+1-xi)} + yi,\ \ i = 1$ then we get y =0.96, and by the same way we can getting the results in Table (7) below:

Table (7): results of defuzzification

X	Belonging degree	Y	Belonging degree	Z	Center	Area	Center * Area
L	0.96	H	0.787878788	H	2.258	0.78787879	1.7790303
L	0.96	H	0.787878788	M	2.1	0.78787879	1.6545455
L	0.96	M	0.21212121	H	2.258	0.21212121	0.4789697
M	0.04	H	0.787878788	H	2.258	0.04	0.09032
M	0.04	L	0	M	2.1		
M	0.04	M	0.21212121	H	2.258	0.04	0.09032
H	0	L	0	M	2.1		
H	0	L	0	H	2.258		
H	0	M	0.21212121	M	2.1		
					Sum =	1.8678788	4.0931855

And the defuzzification in Mamdani inference system we can get COG; $COG = \frac{4.0931855}{1.8678788} =$ 2.1913549744**,** we note that COG is nearest of 2.26 in the results of Z .

Conclusions

The data adopted in this paper is from real spatial data of groundwater levels from Mosul city in Iraq. to get the best estimate of the fitting model. Prediction or so-called kriging techniques require programs for distance matrix, mathematical formulas of prediction, and error variance. The fuzzy concepts and tools preparation rely on the probability theory, so the fuzzy model alone is not sufficient except for its integration and linking with the methods of spatial interpolation in order to expand the prediction process. Two methods of fuzzy inference method and spatial interpolation are characterized by uncertain information that appears in spatial variance with there are common criteria indicating the analysis of spatial distribution. The center of mass takes the output distribution found and its center to outcome the crisp, this is computed as COG in the Mamdani inference system. the results obtained show the convergence between the fitting model and the output graphics in a fuzzy system, which has the same properties in the case of spatial prediction.

References

[1] F. Topaloğlu, & H. Pehlıvan, (2018). Comparison of Mamdani type and Sugeno type fuzzy inference systems in wind power plant installations. *In 2018 6th International Symposium on Digital Forensic and Security* 4–1

[2] M. Sunjana, (2019). Prediction of Production Using the Fuzzy Mamdani Inference System. *International Journal of Advanced Science and Technology (IJAST)*, *28*(6), 136–139.

[3] S. O. Kassim, A. G. Ali, & I. M. Harram, (2021). Design And Implementation Of Mamdani Type Fuzzy Inference System Based Water Level Controller. . *IOSR Journal of Electronics and Communication Engineering*, *16*(4), 15–22.

[4] G. Matheron, (n.d.). *(1973). The intrinsic random functions and their applications. Advances in applied probability* . . 5 (3 ,) 439 – 468 .

[5] H. Wackernagel, (2003). *Multivariate geostatistics: an introduction with applications.* Springer Science & Business Media.

[6] H. Xie, X. Zhang, B.Yu, & H. Sharif, (2011). Performance evaluation of interpolation methods for incorporating rain gauge measurements into NEXRAD precipitation data: a case study in the Upper Guadalupe River Basin. *Hydrological Processes, 25*(24), 3711–3720

[7] A. Jain, & A. Sharma, (2020). *Membership Function Formulation Methods for Fuzzy Logic Systems: a Comprehensive Review*. *7*(19), 2020.

[8] M. Verma, N. K., Singh, V. Rajurkar, & Aqib. (2019). Fuzzy inference network with mamdani fuzzy inference system. In Computational Intelligence. *Theories, Applications and Future Directions," Springer, Singapore.,* 375–388.

[9] L. J. Muhammad, & E. A. Algehyne, (2021). Fuzzy based expert system for diagnosis of coronary artery disease in Nigeria. *Health and Technology*, *11*(2), 319–329.

[10] A. A. Eldeiry, & L. A. Garcia, (2010). Comparison of ordinary kriging, regression kriging, and cokriging techniques to estimate soil salinity using LANDSAT images, *Irrigation and Drainage Engineering, 136*(6), 355–364.

[11] R. Giraldo, L. Herrera, and V. Leiva, (2020). Cokriging prediction using as secondary variable a functional random field with application in environmental pollution. *Mathematics*, *8*(8), 1305.

[12] J.P. Chiles, and P. Delfiner, (2009). *Geostatistics: modeling spatial uncertainty*. John Wiley & Sons.

[13] H. A. Farag, M. A. Gouda, & M. A. Yassin, (2020). The Effect of Using Cokriging against Ordinary Kriging When Estimating Phosphate Deposits at Elsebaiya Area, Egypt. *World Journal of Engineering and Technology*711 ,)4(8 , .

[14] M. Maleki, X. Emery, and N. Mery, (2017). Indicator variograms as an aid for geological interpretation and modeling of ore deposits. *Minerals*, *7*(12), 241.

[15] W. Somayasa, D.K. Sutiari, and W. Sutisna, (2021). Optimal prediction in isotropic spatial process under spherical type variogram model with application to corn plant data. *In Journal of Physics: Conference Series,* 012003.