

## **BALANCING SOIL PARAMETERS AND FARMERS BUDGET BY FEATURE SELECTION AND ORDERED WEIGHTED AVERAGING**

*Marzieh MOKARRAM<sup>1</sup>, Mehran SHAYGAN<sup>2\*</sup>, George Ch. MILIAREISIS<sup>3</sup>*

DOI: 10.21163/GT\_2018.131.08

### **ABSTRACT:**

A method is presented allowing farmers at Shiraz in the Fars province of Iran to balance in between their budget and the soil parameters. First, the three alternatives (Best-First, Greedy-Stepwise and Ranker) of the Feature Selection Method identify the most critical soil fertility parameters. Training data model evaluation indicate that the Greedy-Stepwise feature selection algorithm (with attribute evaluator of CFS-Subset-Eval) presents the highest accuracy for the particular study area. Soil fertility is found to highly depends on Potassium, Phosphor, and Organic Carbon while Copper, Iron, Manganese, and Zinc dependencies are rejected. Finally, by utilizing Ordered Weighted Averaging, six maps with different risk levels in terms of the soil fertility are constructed allowing alternative management options according to the farmers budget. The major scientific contributions are summarized to a) the identification of soil fertility parameters, and the b) construction of maps modeling soil fertility for various degrees of uncertainty allowing agricultural cost effective planning in the study area.

*Key words:* Ordered weighted averaging (OWA); feature selection algorithm; fuzzy; Soil fertility.

## **1. INTRODUCTION**

Crop nutrition and soil fertility elevations are important for sustainable productivity in agricultural lands. There are different methods for determination of soil fertility such as Xie et al. (2015) used principal component analysis for elevation of soil fertility. The results showed that the succession of rocky desertification (RD) had different impacts on soil fertility indicators. Nagaraja et al. (2016) used Generalized soil mass (GSM), Bulk-density-based soil mass (BDSM) and Fine-earth-volume-based (FEV) method for estimations of soil fertility in physically degraded agricultural soils. The results show that the best method for prediction of soil fertility is GSM and BDSM methods. One of the methods for determination of soil fertility is multi-criteria evaluation. The multi-criteria evaluation may be used to develop and evaluate alternative plans which may facilitate a compromise between interested parties (Malczewski, 1996).

Incorporation Multi Criteria Decision Analysis (MCDA) methods and GIS makes a strong tool for spatial planning (Asproth et al., 1999; Belkhiri et al., 2011; Haidu, 2016; Makropoulos et al., 2003; Malczewski & Rinner, 2005; Mohammad et al., 2013).

---

<sup>1</sup> *Department of Range and Watershed Management, College of Agriculture and Natural Resources of Darab, Shiraz University, Iran, Email: m.mokarram@shirazu.ac.ir*

<sup>\*2</sup> *K. N. Toosi University of Technology, 19967-15433 Tehran, Iran, Email: mehranshaygan@gmail.com\*

<sup>3</sup> *Open University of Cyprus, 38, Tripoleos Str., 104-42 Athens, Greece, e-mail: miliareisis.g@gmail.com*

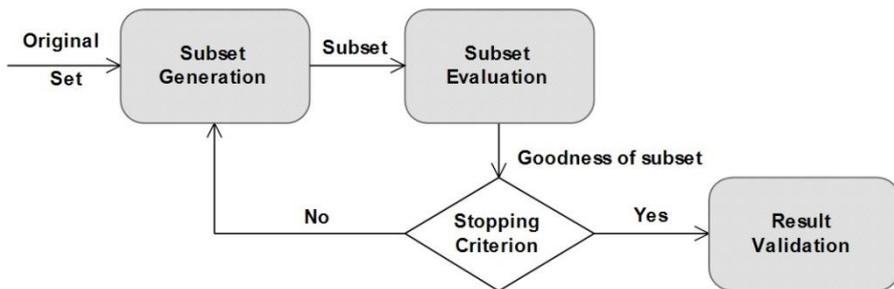
The aim of the present study is to prepare the soil fertility maps based on the OWA operators of GIS-based multi-criteria evaluation procedures and feature selection west of Shiraz city, in the Fars province that is one of the most important centers of agriculture in Iran. Such a research effort is expected to provide the framework for cost effective agricultural planning in the particular study area. First, the factors determining soil fertility in the study area should be quantified, and then OWA method is expected to specify soil fertility at different risk levels.

## 2. METHODOLOGY

In order to prepare the soil fertility maps training data at 45 test sites were collected. Then Feature Selection Algorithm is expected to identify the most significant soils properties specifying soil fertility. Then maps will be constructed by spatial interpolation per significant soil property while fuzzy parameter maps should allow the definition of different risk levels with OWA. The methods description follows here under.

### 2.1. Feature selection

In order to select the most significant soil properties for soil fertility, the feature selection method is implemented. Feature selection has four steps (**Fig. 1**): a) Generation procedure, b) Evaluation subset, c) Stopping criterion and d) Validation procedure. The feature selection implementation is shown in **Fig. 1**.



**Fig. 1.** The feature selection implementation.

Weka v.3.8 as learning machine (Waikato University, 2015) is used for Feature Selection. Three search methods which include Best-First, GreedyStepwis and Ranker as search method are tested as attribute evaluators. More specifically: a)for Best-First method, the CFS-Subset-Eval is used, b) For Greedy-Stepwise method, the CFS-Subset-Eval is used, and c)for Ranker method, the Info-Gain-Attribute-Eval, Gain-Ratio-Attribute-Eval, Symmetricer-Attribute-Eval, Relife-FAttribute-Eval, and Principal-Components are used.

**Best-First:** In the search method using Greedyhill Climbing augmented with a backtracking facility performs space of attribute subsets. In the search method starts with the empty set or full set of attributes and search forward or backward respectively. Also it start at any point and search in two directions.

**GreedyStepwis:** Using the space of attribute subsets performs a greedy forward or backward search. The search method starts with no, all attributes or from a free point in the space. **Ranker:** It ranks attributes by their individual evaluations.

In previous research efforts (Dash & Liu, 2003; Naseriparsa et al., 2013) the performance evaluation of the Feature Selection Algorithm is based on a) the Average Number of Misclassified Samples (AMS, see equation 1) and b) on the Average Relative Absolute Error (ARAE, see equation 2).

$$AMS = \frac{\sum_{i=1}^n MS_i}{N} \tag{1}$$

Where

MS<sub>i</sub> is the number of misclassified for each models and N is the sampling points.

$$ARAE = \frac{\sum_{i=1}^n RAE_i}{N} \tag{2}$$

Where

RAE<sub>i</sub> is the relative absolute error for the classification models and N is the sampling points.

### 2.2. Inverse Distance Weighted (IDW)

The IDW method (Burrough & McDonnell, 1998) was used for interpolating the effective data in order to determine the soil fertility at regular grid. The IDW method (see equation 3) is actually a distance-weighted average of the sampled points at a defined neighborhood.

$$\hat{z}(x_0) = \frac{\sum_{i=1}^n z(x_i) d_{ij}^{-r}}{\sum_{i=1}^n d_{ij}^{-r}} \tag{3}$$

Where  $x_0$  is the estimation point and  $x_i$  are the data points within a chosen neighborhood. Weights ( $r$ ) are related to the distance by  $d_{ij}$ .

### 2.3. Ordered Weight Average (OWA)

Based on the input data, the OWA combination operator associates with the  $i$ -th location of a set of order weights  $v = v_1, v_2, \dots, v_n$  such that  $v_j \in [0, 1], j=1,2,\dots,n$ ,

$\sum_{j=1}^n v_j = 1$ , and it is defined in equation 4 (Malczewski et al., 2003).

$$OWA_i = \sum_{j=1}^n \left( \frac{u_j v_j}{\sum_{j=1}^n u_j v_j} \right) z_{ij} \tag{4}$$

Where  $z_{i1} \geq z_{i2} \geq \dots \geq z_{in}$  are the sequences obtained by reordering the attribute values  $a_{i1}, a_{i2}, \dots, a_{in}$ , and  $u_j$  is the criterion weight reordered based on the attribute value,  $z_{ij}$ .

### 2.4. Study Area

The study area is located at latitude of N 29° 34' - 29° 36' and longitude of E 52° 49' to 52° 57' (Fig. 2) in the west of Shiraz, Iran. It is an area of about 44 km<sup>2</sup>. The elevation of the study area ranges from 1,574 m to 1,722 m.

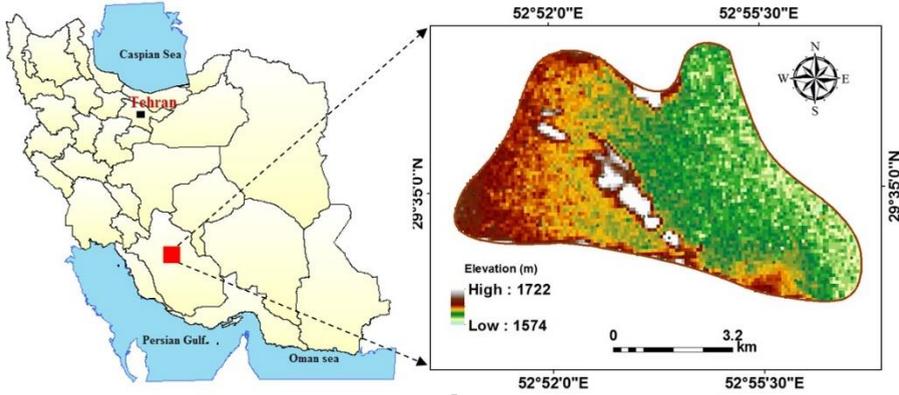


Fig. 2. Study Area and its location

The case study has a mid-latitude semi-arid cool climate (Köppen-Geiger classification: BSk). Based on the Holdridge life zones system of bioclimatic classification Shiraz is situated in or near the warm temperate thorn steppe biome. Details about climate of the study area such as temperature, precipitation are shown in Fig. 3.

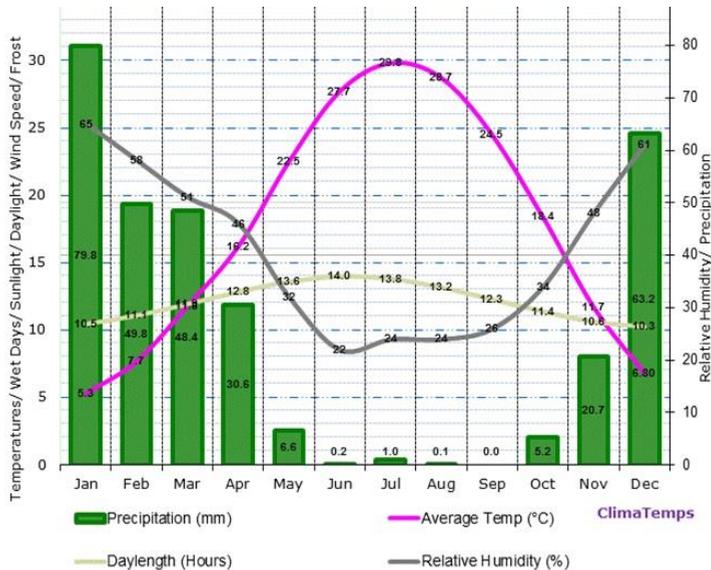


Fig. 3. Details about climate of the study area such as temperature, precipitation

### 3. RESULTS

In order to predict the variability of the soil fertility, some minerals were used which are named here as potassium (K), phosphor (P), copper (Cu), iron (Fe), manganese (Mn), organic carbon (OC) and zinc (Zn) (0-60 cm of soil surface); then, maps of each parameter were prepared (**Table 1**) (Organization of Agriculture, Jihad Fars province). In continue using feature selection algorithm, IDW, Fuzzy method, AHP were determined soil fertility maps with different risk levels.

**Table 1.**

**Descriptive statistics of the data for the soil fertility  
(Organization of Agriculture, Jihad Fars province)**

Statistic parameters	K (mg/kg)	P (mg/kg)	Cu (mg/kg)	Fe (mg/kg)	Mn (mg/kg)	OC (mg/kg)	Zn (mg/kg)
maximum	666.00	30.00	2.00	15.00	52.50	1.65	3.00
minimum	137.00	2.00	0.20	1.00	2.80	0.18	0.10
average	313.73	13.94	0.97	4.54	14.77	1.01	0.65
STDEV	104.28	6.49	0.36	2.84	10.71	0.35	0.50

#### 3.1. Feature selection

In the study for selection of importance data using feature selection algorithm was used Weka v.3.8. In the study different combination of feature selection method are used such as: Best-First, GreedyStepwis and Ranker as search method. For Best-First used CFS-Subset-Eval as attribute evaluator. For Greedy-Stepwise used CFS-Subset-Eval as attribute evaluator. While Info-Gain-Attribute-Eval, Gain-Ratio-Attribute-Eval, Symmetricer-Attribute-Eval, Relife-FAttribute-Eval, Principal-Components used as attribute evaluator for ranker method. The more details of each methods with attribute evaluator show in **Table 2** and **Table 3**. According to **Table 2** the importance data for determination of soil fertility for Best-First (CFS-Subset-Eval) and Greedy-Stepwise (CFS-Subset-Eval) were Fe, Mn and Cu. Cu, Fe and Mn were as the importance data for Ranker (Gain-Ratio-Attribute-Eval). For Ranker (Symmetricer-Attribute-Eval), N, P and K was as importance data. For Ranker (Relife-FAttribute-Eval) the best data for determination of soil fertility were Cu, Fe and P. while for Ranker (Principal-Components), Cu, P and Mn was as importance data

**Table 2.**

**Features selected by different feature selection methods**

Search method	Attribute evaluator	Selected Features
Best-First	CFS-Subset-Eval	Fe, Mn, Cu
Greedy-Stepwise	CFS-Subset-Eval	Fe, Mn, Cu
	Info-Gain-Attribute-Eval	Cu, Fe, Mn
Ranker	Gain-Ratio-Attribute-Eval	Mn, P, Fe
	Symmetricer-Attribute-Eval	OC, P, K
	Relife-FAttribute-Eval	Cu, Fe, P
	Principal-Components	Cu, P, Mn

For analysis of feature selection algorithm and select the best methods and their performance are evaluated using C4.5 (J48) classifier. Information of the classifier show in **Table 3**. According to **Table 3** determined that the best method with the lowest number of incorrect and the highest correlation of coefficient was Ranker (Relife- squared error FAttribute-Eval).

**Table 3.****Evaluation of classifiers based on J48**

Search method	Attribute evaluator	Classifier model	Test mode	Correlation coefficient	Mean absolute error	Root mean squared error	Relative absolute error	Root relative of error	Number of correct	Correct
Best-First	CFS-Subset-Eval	Full training set	10 Fold class validates	75.29	0.105	0.229	46.77	68.62	64	21
Greedy-Stepwise	CFS-Subset-Eval	Full training set	10 Fold class validates	75.29	0.105	0.229	46.77	68.62	64	21
	Info-Gain-Attribute-Eval	Full training set	10 Fold class validates	75.29	0.105	0.229	46.77	68.62	64	21
	Gain-Ratio-Attribute-Eval	Full training set	10 Fold class validates	75.29	0.105	0.229	46.77	68.62	64	21
Ranker	Symmetric-Attribute-Eval	Full training set	10 Fold class validates	75.29	0.105	0.229	46.77	68.62	64	21
	Relife-FAttribute-Eval	Full training set	10 Fold class validates	88	0.0706	0.15	32.89	57.11	80	5
	Principal-Components	Full training set	10 Fold class validates	75.29	0.105	0.229	46.77	68.62	64	21

Finally for determination of error and performance values used AMS and ARAE. The results of two methods show in **Table 4**. According to **Table 4** Ranker (Relife-FAttribute-Eval) had the lowest AMS (7.69) and ARAE (38.69) that selected as the best method for extraction of the importance data for determination of soil fertility.

**Table 4.****Value of performance**

Search method	Attribute evaluator	Performance method	
		ARAE	AMS
Best-First	CFS-Subset-Eval	55.02	32.31
Greedy-Stepwise	CFS-Subset-Eval	55.02	32.31
	Info-Gain-Attribute-Eval	55.02	32.31
	Gain-Ratio-Attribute-Eval	55.02	32.31
Ranker	Symmetric-Attribute-Eval	55.02	32.31
	Relife-FAttribute-Eval	38.69	7.69
	Principal-Components	55.02	32.31

Finally OC, P and K from Ranker (Relife-FAttribute-Eval) was used for determination of soil fertility in OWA method.

### 3.2. Inverse Distance Weighted (IDW)

The IDW interpolation was used to predict K, P and OC values which are all shown in Fig. 4. According to Fig. 4, within the chosen study area, most elements in the north and parts of the south were determined to have lower amounts than other regions.

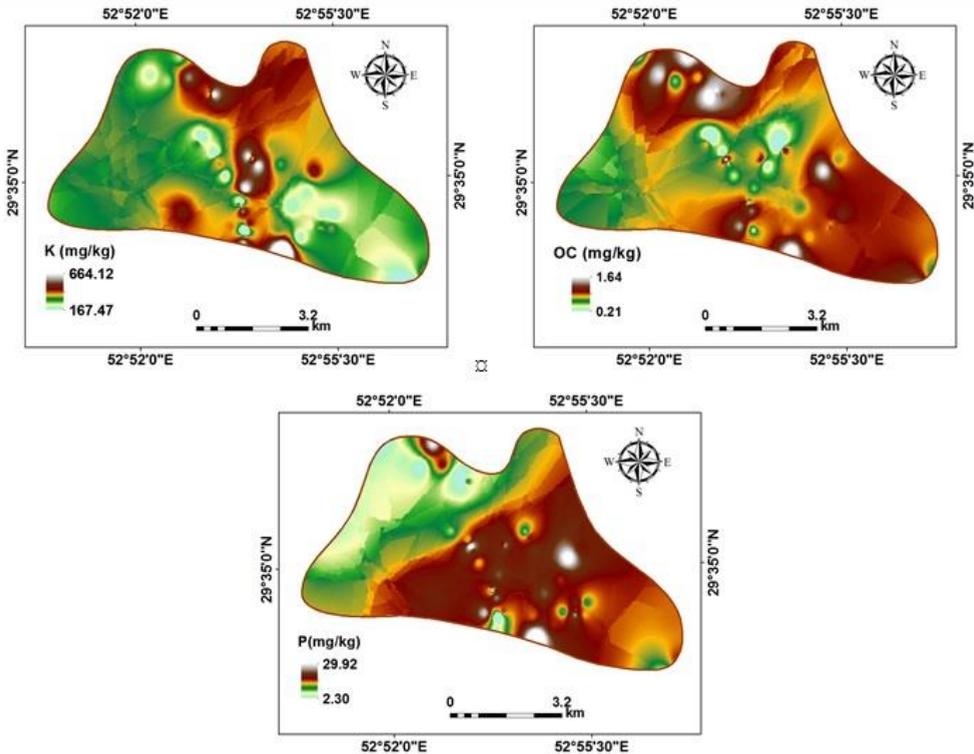


Fig. 4. The map of K, P and OC values which are interpolated by IDW

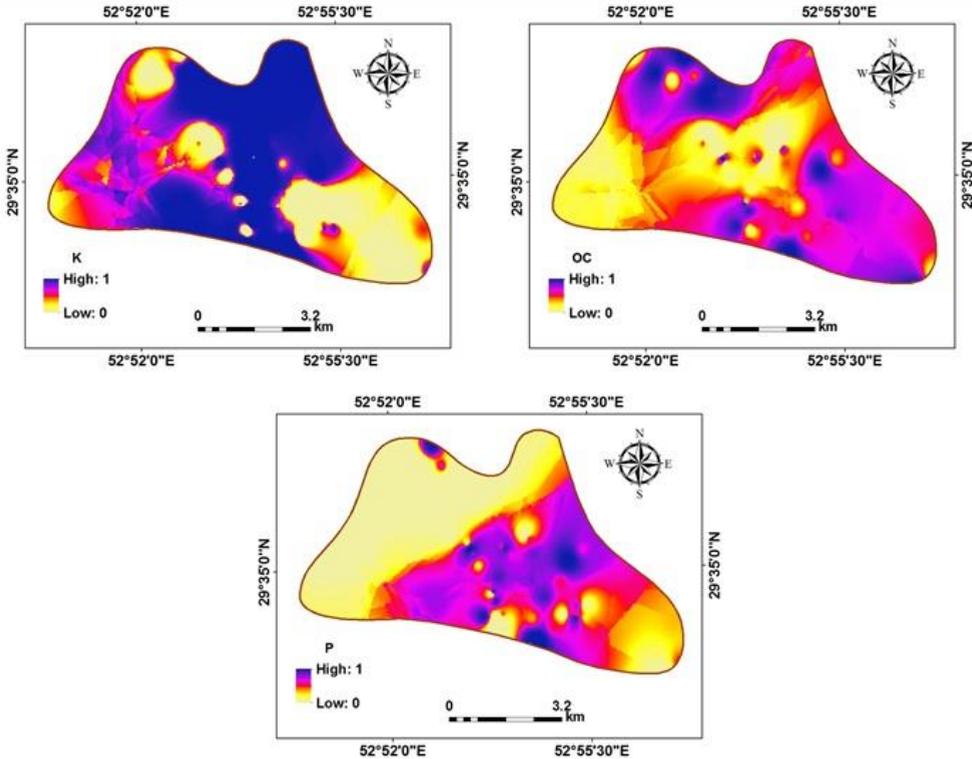
### 3.3. Fuzzy method

The membership function for each parameter (K, P and OC) was defined according to FAO (1983), and each fuzzy map was created with value range in between 0 and 1 (Fig. 5).

According to Fig. 5, the mainly east and southeast of the study area had suitable values for P parameter with fuzzy membership value close to 1, an exception being the parts in center and west of the study area. Also according to the K fuzzy map, some parts of the north, southeast and west were not suitable as well.

### 3.4. OWA method

In order to use OWA method was used IDRISI32 software. **Table 5** consist of six generic sets of order weights for seven factors: (1) an average level of the risk and a full trade-off, (2) a low level of the risk and no trade-off, (3) a high level of the risk and no trade-off, (4) a low level of the risk and an average trade-off, (5) a high level of the risk and an average trade-off, (6) an average level of the risk and no trade-off (**Table 5**).



**Fig. 5.** Fuzzy Map for soil parameters including k, P and OC

According to **Fig. 6**, with decreasing risk (no trade-off) (**Fig. 6 (2)**), the area with high soil fertility was determined. So, that only the parts of center, south and southeast of the study area was suitable for soil fertility. Also with increasing risk (no trade-off) (**Fig. 6 (3)**) all of the study had good soil fertility. According to **Fig. 6 (3)** almost the all of the study area had high soil fertility. With average risk (full trade-off) (**Fig. 6 (1)**) the all of effective parameters of soil fertility were received some weight (0.33). According to **Fig. 6 (1)** the parts of the study area had good value (except a parts of center and southwest of the study area). The **Fig. 6 (4)** showed low risk with average trade-off that in comparison of **Fig. 6 (2)** had more risk. The **Fig. 6 (5)** showed high risk with average trade-off that in comparison of **Fig. 6 (3)** had lower risk for determination of soil fertility. **Fig. 6 (6)** showed average risk with no trade-off that in comparison of **Fig. 6 (3)** had more risk.

Table 5.

Typical sets of order weights for three factors.

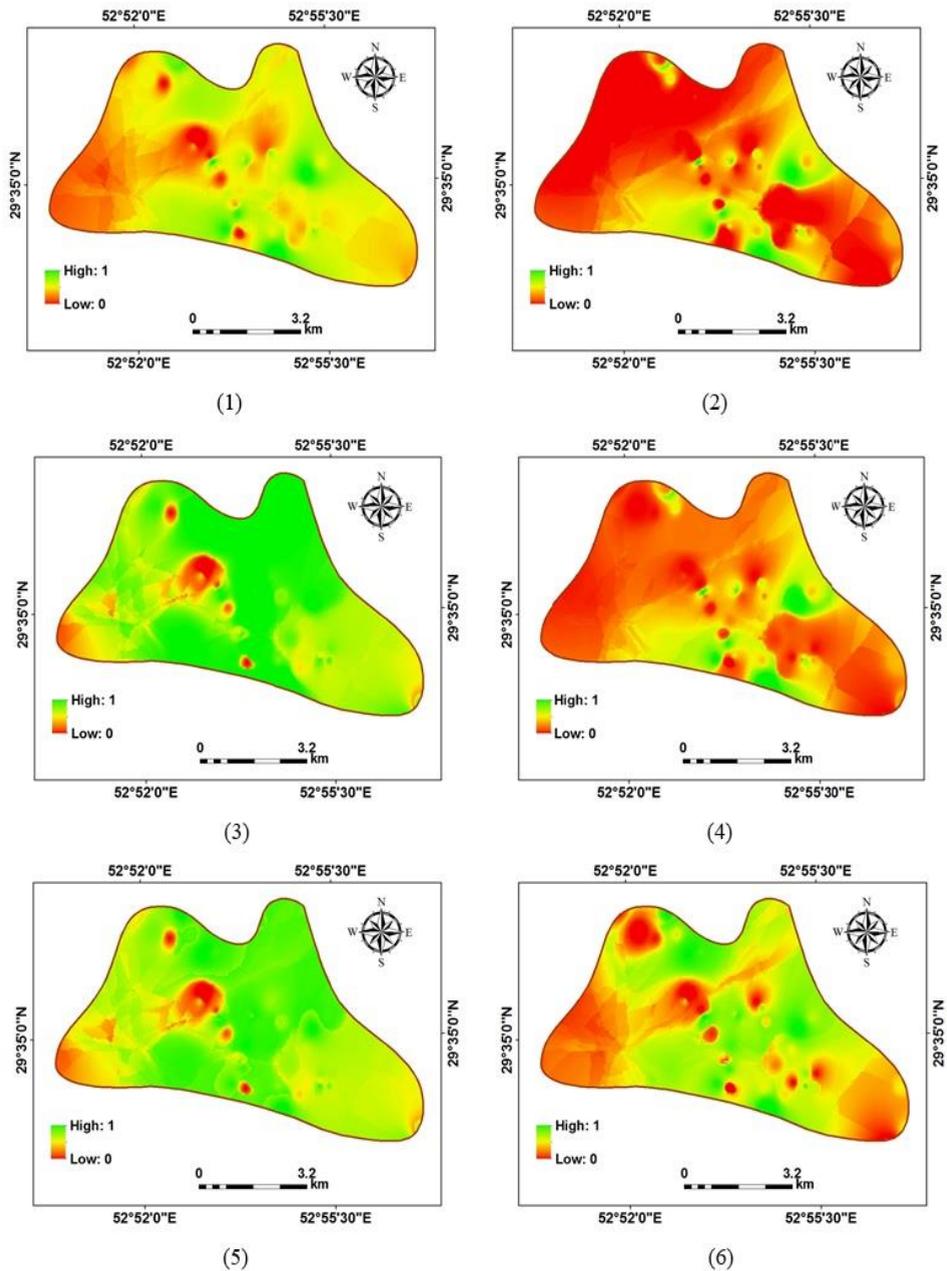
(1) Average level of risk and full trade-off			
order weight	0.33	0.33	0.33
rank	1st	2nd	3rd
(2) Low level of risk and no trade-off			
order weight	1	0	0
rank	1st	2nd	3rd
(3) High level of risk and no trade-off			
order weight	0	0	1
rank	1st	2nd	3rd
(4) Low level of risk and average trade-off			
order weight	0.8	0.2	0
rank	1st	2nd	3rd
(5) High level of risk and average trade-off			
order weight	0	0.2	0.8
rank	1st	2nd	3rd
(6) Average level of risk and no trade-off			
order weight	0	1	0
rank	1st	2nd	3rd

In fact using feature selection and select the importance data and then OWA method not only can prepare soil maps with multi-criteria decision, but also save time and money in soil science.

One of the main advantages of OWA is its ability in generating the wide variety of possibilities for different designs. In recent years some studies about soil fertility using OWA method is done by Delsouz Khaki et al. (2015), Bijanzadeh and Mokarram (2013) and Mokarram and Bardideh (2012). In some of them soil fertility is achieved by combining fuzzy algorithm and OWA method. In current research only a medium risk was used and the different risk levels is not evaluated. It is obvious that the OWA method with different risk levels can help a user such as farmer to make different decisions based on different financial situations and risk levels.

#### 4. CONCLUSIONS

The aim of the present study was to produce the soil fertility maps based on feature selection algorithm and OWA operators of the GIS-based multi-criteria evaluation procedures in the west of Shiraz city of Fars province (Iran). Using the feature selection algorithm (Best-First, Greedy-Stepwise and Ranker) was selected the most importance parameters. Then using the OWA approach was provided a mechanism for guiding the decision maker/analyst through the multi-criteria combination procedures.



**Fig. 6.** The OWA map for study area based on different trade-offs. (1) an average level of the risk and a full trade-off, (2) a low level of the risk and no trade-off, (3) a high level of the risk and no trade-off, (4) a low level of the risk and an average trade-off, (5) a high level of the risk and an average trade-off, (6) an average level of the risk and no trade-off

Results showed that with decreasing the risk (no trade-off), the area with a high soil fertility was determined. Therefore, just parts of the east and southeast of the study area were considered suitable for the soil fertility. Furthermore, with increasing the risk (no trade-off), almost all of the study area had a good soil fertility. In fact using feature selection and select the importance data and then OWA method not only can prepare soil maps with multi-criteria decision, but also save time and money in soil science.

## ACKNOWLEDGEMENTS

The authors would like to thanks to all personnel of Natural Resources Organization of Fars province and Organization of Agriculture Jihad Fars for their kind help.

## REFERENCES

- Asproth, V., Holm berg, S.C. & Håkansson, A. (1999) *Decision support for spatial planning and management of human settlements*, in Lasker, G.E. (Ed.): *Advances in Support Systems Research*, Vol. 5, International Institute for Advanced Studies in Systems Research and Cybernetics, Windsor, Ontario, Canada, 30–39.
- Belkhiri, L., Boudoukha, A. & Mouni, L. (2011) A multivariate Statistical Analysis of Groundwater Chemistry Data, *International Journal of Environmental Research*, 5 (2), 537-544.
- Bijanzadeh, E. & Mokarram, M. (2013) The use of fuzzy- AHP methods to assess fertility classes for wheat and its relationship with soil salinity: east of Shiraz, Iran : A case study. *Australium journal of crop science*, 7(11), 1699-1706.
- Burrough, P.A. & McDonnell, R.A. (1998) *Principles of geographical information systems. Spatial Information System and Geostatistics*. Oxford University Press, New York.
- Dash, M. & Liu, H. (2003) Consistency-based Search in Feature Selection, *Artificial Intelligence*, 151, 155-176.
- Delsouz Khaki, N., Honarjoo, N., Davatgar, A. & Torab, H. (2015) Soil Fertility Evaluation Using Fuzzy Membership Function (Case Study: Southern Half of Foumanat Plain in North of Iran). *Allgemeine forst undjagdzeitung*, 53-64.
- Haidu, I. (2016). What Is Technical Geography. *Geographia Technica*, 11(1), 1-5. DOI: 10.21163/GT\_2016.111.01
- Makropoulos, C., Butler, D. & Maksimovic, C. (2003) A fuzzy logic spatial decision support system for urban water management. *Jornal of Water Resource. Planning. Manage*, 129 (1), 69–77.
- Malczewski, J.A. (1996) GIS-based approach to multiplecriteria group decision making. *International Journal of Geographical Information Systems*, 10(8), 955-971.
- Malczewski, J., Chapman, T., Flegel, C., Walters, D., Shrubsole, D. & Healy, M.A. (2003) GIS-multicriteria evaluation with ordered weighted averaging (OWA): case study of developing watershed management strategies. *Environment Planning, A* 35 (10), 1769–1784.

- Malczewski, J. & Rinner, C. (2005) Exploring multicriteria decision strategies in GIS with linguistic quantifiers: a case study of residential quality evaluation. *Journal of Geography System*, 7(2), 249–268.
- Mohammad, M., Sahebgharani, A. & Malekipour, E. (2013). Urban growth simulation through cellular automata (CA), analytic hierarchy process (AHP) and GIS; case study of 8th and 12th municipal districts of Isfahan. *Geographia Technica*, 8(2), 57-70.
- Mohammad, M., Sahebgharani, A. & Malekipour, E. (2013). Urban growth simulation through cellular automata (CA), analytic hierarchy process (AHP) and GIS; case study of 8th and 12th municipal districts of Isfahan. *Geographia Technica*, 8(2), 57-70.
- Mokarram, M. & Bardideh M. (2012) Soil fertility evaluation for wheat cultivation by fuzzy theory approache and compared with boolean method and soil test method in gis area. *Agronomy journal (pajouhesh & sazandegi)*, 25(3), 111 -123.
- Nagaraja, M. S., Bhardwaj, A. K., Reddy, G. V. P., Srinivasamurthy, C. A., and Kumar, S. (2016) Estimations of soil fertility in physically degraded agricultural soils through selective accounting of fine earth and gravel fractions. *Solid Earth*, 7, 897-903, <https://doi.org/10.5194/se-7-897-2016>.
- Naseriparsa, M., Bidgoli, A.M. & Varaee, T. (2013) A Hybrid Feature Selection method to improve performance of a group of classification algorithms. *International Journal of Computer Applications*, Vol 69, No 17, 28-35.
- Waikato University (2015) Weka software. Weka for Windows, v. 3.8.
- Xie, L., W, Zhong, J., Chen, F.F., Cao, F.X., Li, J.J. & Wu, L.C.(2015) Evaluation of soil fertility in the succession of karst rocky desertification using principal component analysis, *Solid Earth*, 6, 515-524.