

Vol. 20

Original Research Paper

di https://doi.org/10.46488/NEPT.2021.v20i04.047

Open Access Journal

2021

Assessment of Soil Environment Pollution Based on Fuzzy Pattern Recognition Model

Xin Huang[†] and Lin Qiu

College of Water Resources, North China University of Water Resources and Electric Power, Zhengzhou, Henan, 450046 China

†Corresponding author: Xin Huang; huangxin0013@163.com

Nat. Env. & Poll. Tech. Website: www.neptjournal.com

Received: 16-11-2020 Revised: 21-12-2020 Accepted: 22-02-2021

Key Words: Soil environment Pollution Fuzzy pattern recognition General weighted length Membership degree assessment

ABSTRACT

As one of the basic elements of the ecosystem and natural environment, the soil is closely related to human life. However, the problem of soil environment pollution is becoming more and more serious, which needs to be solved urgently. It will provide a reference for solving the problem of soil environment pollution if a suitable method can be found to evaluate the degree of soil environment pollution. The degree of soil environment pollution belongs to the fuzzy concept. It is a fuzzy pattern recognition problem to evaluate the degree of soil environment pollution according to the soil environment level standard value. The fuzzy pattern recognition method of soil environment pollution assessment makes full use of the fuzzy characteristics existing in soil environment assessment, considers the common influence of the weight of evaluation factors and the index value, and also considers the correlation between indexes and the similarity of index characteristics of different sampling points, therefore, the interference of some man-made certain factors is concealed. The model was applied to the soil sampling point in Wudi County of Shandong Province in the Yellow River basin of China. The results showed that the evaluation of soil environment pollution by this method accorded with the actual situation and can provide a basis for ensuring the stability of soil ecological environment, improving the quality of cultivated land, and improving regional ecological conditions. At the same time, in view of the current problem of soil environment pollution, the paper calls on human beings to pay attention to environmental protection and be responsible for their own safety and also proposes four aspects of treatment schemes and measures.

INTRODUCTION

Soil is one of the basic elements of the ecosystem. It plays an important role in maintaining biodiversity, stabilizing and buffering environmental changes in terrestrial ecosystems. It is a component of the natural environment, connecting the atmosphere, hydrosphere, lithosphere, and biosphere (Newman et al. 2020, Mammo et al. 2019, Ren et al. 2016). It is an important guarantee for promoting sustainable development of the economy and society. Soil is closely related to human life, and the quality of the soil environment has a great impact on human beings. Soil pollution refers to the phenomena in which various toxic and harmful chemicals penetrate the soil and cause changes in soil characteristics, which have a negative impact on plants and animals. Longterm, irreversible, hysteresis, and concealment are all characteristics of soil pollution. Toxic and harmful substances entering the soil environment will have a great negative impact on the structure, composition, and function of the soil, resulting in a significant decline in soil quality, poor soil nutrients, and soil composition necrosis, affecting food security, ecological security, the physical health of residents (Francisca & Cristina 2019), etc., which will have a great impact on agricultural development and seriously hinder economic development. At present, researches on soil environment pollution mainly focus on the spatial distribution of pollutants, regional pollution characteristics, present situation investigation, soil remediation, and evaluation, etc. There are many methods for soil environment pollution assessment, such as multi-factor comprehensive assessment, Nemerow pollution index assessment, principal component analysis, etc. Among them, the multi-factor comprehensive assessment has wide adaptability and simple application, but the standard boundary is too clear (Su et al. 2016). The mathematical process of the Nemero pollution index assessment is simple, but it only considers pollution factors with the highest concentrations and excludes pollution factors with low concentrations but high damage. (Kong et al. 2019). Human factors influence the limit value of the load coefficient chosen from each eigenvector by principal component analysis, and the weight changes when the variables are categorized into various principal components, making the

significance test of evaluation factors impossible (Onwosi et al. 2020 Because the evaluation findings of each single soil environment factor are often incompatible, direct use of a relevant standard critical value to divide the level of soil environment pollution cannot produce an exact evaluation in the study of soil environment pollution assessment

The quality of the soil environment is determined by the factors constituting the soil environment. Before the evaluation of soil environment pollution, the advantages and disadvantages of each evaluation factor should be evaluated first. The evaluation factors are gradually changing between the advantages and disadvantages, and show the doublesided property in the intermediate transition. In other words, the level of soil environment pollution is a fuzzy concept, and evaluating the level of soil environment pollution according to the standard level value of soil environment pollution is a fuzzy pattern recognition problem (Poomagal et al. 2020). Therefore, this paper explores the application of the principle and method of fuzzy pattern recognition in the assessment of soil environment pollution and expounds in detail the mathematical and physical significance of the recognition criterion with the minimum sum of squares of weighted distances and the corresponding iterative formula.

MATERIALS AND METHODS

Comprehensive Evaluation Theory and Model of Fuzzy Pattern Recognition

Let a sample set consist of *n* soil samples, and each sample has *m* evaluation indexes. Then, the measured index matrix $X_{m \times n} = (x_{ij})$ is obtained, where $i = 1, 2, ..., m; j = 1, 2, ..., n, x_{ij}$ is the measured value of index *i* in the sample *j*. The *m* indexes are evaluated according to level *c* soil standard, then the index standard matrix $Y_{m \times c} = (y_{ih})$ is obtained, where h = 1, 2, ..., c; y_{ih} is the level *h* standard value of index *i*.

To describe the fuzziness of soil environment pollution, the matrix $X_{m\times n}$ and $Y_{m\times c}$ were converted into the relative membership degree matrix of measured indexes $R_{m\times n} = (r_{ij})$ and the relative membership degree matrix of index standards $S_{m\times c} = (s_{ih})$ by using the relative membership degree. It is stipulated here that: for the index *i* that the higher the measured concentration, the heavier the pollution, when $x_{ij} \le y_{i1}$, its relative membership degree for pollution $r_{ij} =$ 0; when $x_{ij} \ge y_{ic}$, $r_{ij} = 1$. For the index *i* that the higher the measured concentration, the lighter the pollution, when $x_{ij} \ge y_{i1}$, its relative membership degree for pollution $r_{ij} = 0$; when $x_{ij} \le y_{ic}$, $r_{ij} = 1$; and when x_{ij} between y_{i1} and y_{ic} , r_{ij} is determined according to the linear interpolation formula $r_{ij} =$ $(x_{ij^-}y_{i1})/(y_{ic}-y_{i1})$. It can also be stipulated that the relative membership degree to pollution of the level 1 standard value of index *i* is 0; the relative membership degree to pollution of level *c* standard value of index *i* is 1. The relative membership degree s_{ih} of level *h* standard value of index *i* which is between level 1 and level *c* was determined by formula $s_{ih}=(y_{ih} - y_{i1})/(y_{ic} - y_{i1})$.

Sample *j* is subordinate to soil environmental index standard of level *h* with relative membership degree u_{hj} , then, the relative membership degree matrix $U_{c\times n}=(u_{hj})$, which satisfies the constraint condition

$$\sum_{h=1}^{c} u_{hj} \text{ -1=0 , } \forall j \text{ ; } \sum_{j=1}^{n} u_{hj} > 0 \text{ , } \forall h$$

The relative membership degree matrix of the measurement index $R_{m\times n}$ not only describes the superscalar but also represents the weight. Here, its normalized matrix $W_{m\times n}=(w_{ij})$ is defined as the superscalar weight matrix, which satisfies

$$\sum_{i=1}^m w_{ij} = 1$$
 , $\forall j_{\circ}$

Next, solve for u_{hj} . The difference between sample *j* and soil environmental index standard of level *h* can be expressed by general weighted length, which is:

$$D(r_{j}, s_{k}) = u_{kj} \sqrt[p]{\sum_{i=1}^{m} (w_{ij} | r_{ij} - s_{ik} |)^{p}} \qquad \dots (1)$$

The optimal relative membership degree of sample j to soil environmental index standard of level h is obtained by taking the minimum sum of squares of general weighted length as the objective function, which is:

$$\min\left\{F(u_{hj}) = u_{hj}^{2} \left[\sum_{i=1}^{m} \left(w_{ij} | r_{ij} - s_{ih}|\right)^{p}\right]^{\frac{2}{p}}\right\} \qquad \dots (2)$$

Where, distance parameter p = 1 is the hamming distance, and p = 2 is the Euclidian distance.

The Lagrange equation is constructed according to the objective function (2) and its constraints, which is:

$$L(u_{hj},\lambda) = u_{hj}^{2} \left[\sum_{i=1}^{m} \left(w_{ij} | r_{ij} - s_{ih} | \right)^{p} \right]^{\frac{2}{p}} - \lambda \left(\sum_{h=1}^{c} u_{hj} - 1 \right) \quad \dots (3)$$

Then, seeking the partial derivatives of equation (3) respectively for u_{hj} and Lagrange multiplier I, setting the partial derivatives equal to 0, we get

$$u_{hj} = \frac{1}{\sum_{k=1}^{c} \left[\frac{\sum_{i=1}^{m} (w_{ij} | r_{ij} - s_{ih} |)^{p}}{\sum_{i=1}^{m} (w_{ij} | r_{ij} - s_{ik} |)^{p}} \right]^{\frac{2}{p}}} \dots (4)$$

When formula (4) is used for evaluation and calculation, the value range of k is dynamic and should be determined by comparing m indexes of sample j in the matrix $R_{m\times n}$ and the matrix $S_{m\times c}$. Let the value of relative membership degree of m indexes of sample j be $r_j = (r_{1j}, r_{2j}, r_{mj})^T$. Compare the relative membership degree of r_{ij} of index i(i = 1, 2, m) one by one with the relative membership degree of $s_i = (s_{i1}, s_{i2}, ..., s_{ic})$ of criteria of index i in $S_{m\times c}$. If the minimum level of index i falling into s_i interval is a_j and the maximum level is b_j , then a_j and b_j are the value range of k, with $1 \le a_j < b_j \le c$. In formula (4), the value range of h is the same as that of k,

obviously when
$$h < a_j$$
 or $h < b_j$, $u_{hj} = 0$; $\sum_{h=a_j}^{b_j} u_{hj} = 1$.

In conclusion, the complete form of the theoretical model of fuzzy pattern recognition and evaluation of soil environmental pollution can be expressed as follows:

Where, $j = 1, 2, ..., n; 1 \le a_j < b_j \le c; h = a_j, a_j + 1, ..., b_j$. When p = 1, 2, the relative membership degree matrix of the sample set to level h is, ${}_1U_{c\times n} = ({}_1u_{hj})$ and ${}_2U_{c\times n} = ({}_2u_{hj})$ respectively. Here, the average relative membership degree matrix of the sample sets of two kinds of distance parameters for level h: $\overline{U}_{c\times n} = \frac{1}{2}({}_1u_{hj} + {}_2u_{hj}) = \overline{u}_{hj}$ is taken

to eliminate the influence of different values of distance

parameter p on the comprehensive evaluation of soil environmental pollution level identification.

According to the average relative membership degree matrix, the eigenvectors of *n* soil samples belonging to each level are:

$$H = (1, 2, ..., c) \cdot (\overline{u}_{hj}) = (H_1, H_2, ..., H_n) \qquad ...(6)$$

According to the level characteristic values of samples, the classification of soil environmental pollution was comprehensively evaluated (Chen 1998).

RESULTS AND DISCUSSION

Example Application

After the soil environment is polluted, it will have a great negative impact on the structure, composition, and function of soil, leading to the decrease of soil nutrients and soil hardening, resulting in the impoverishment of soil nutrients and the necrosis of soil components. Therefore, the evaluation of nutrient level in soil environment can reflect the degree of soil environment pollution, that is, the higher the nutrient level of soil environment is, the lighter the pollution is; while the lower the nutrient level of soil environmental is, the more serious the pollution is.

In this paper, partial results of soil environmental nutrient determination in the sampling point in reference (Lv 2018) (the sampling point is located in Wudi County, Binzhou City, Shandong Province, China, in the Yellow River Basin) were selected as the basic data for evaluating soil environment nutrient levels, as shown in Table 1.

The second soil survey classification standard in China used the standard concentration values of the soil environment nutrient level evaluation index. Table 2 shows the selected critical values of the evaluation factor level index based on the circumstances of soil sample determination.

According to Table 1 and Table 2:

Sampling point number	organic matter (%)	total nitrogen (%)	available nitrogen (mg/kg)	available phosphorus (mg/kg)	rapidly available potassium (mg/kg)
1	0.95	0.085	45	8.5	98
2	1.28	0.071	76	3.9	196
3	2.67	0.274	120	13.6	137
4	1.89	0.937	53	7.7	185
5	6.74	1.136	95	35	230
6	3.86	0.168	104	18	149

Table 1: Determination results of physicochemical characters of soil environment nutrient level.

level	Ι	II	III	IV	V
organic matter (%) (>)	4.0	3.0	2.0	1.0	0.6
total nitrogen (%) (>)	0.20	0.15	0.1	0.075	0.05
available nitrogen (mg/kg) (>)	150	120	90	60	30
available phosphorus(mg/kg) (>)	40	20	10	5	3
rapidly available potassium (mg/kg) (>)	200	150	100	50	30

Table 2: Critical values of the level indexes of each evaluation factor.

$$X_{5\times6} = \begin{bmatrix} 0.95 & 1.28 & 2.67 & 1.89 & 6.74 & 3.86 \\ 0.085 & 0.071 & 0.274 & 0.937 & 1.136 & 0.168 \\ 45 & 76 & 120 & 53 & 95 & 104 \\ 8.5 & 3.9 & 13.6 & 7.7 & 35 & 18 \\ 98 & 196 & 137 & 185 & 230 & 149 \end{bmatrix} = (x_{ij}),$$
$$Y_{5\times5} = \begin{bmatrix} 4.0 & 3.0 & 2.0 & 1.0 & 0.6 \\ 0.20 & 0.15 & 0.1 & 0.075 & 0.05 \\ 150 & 120 & 90 & 60 & 30 \\ 40 & 20 & 10 & 5 & 3 \\ 200 & 150 & 100 & 50 & 30 \end{bmatrix} = (y_{ih})$$

where, *i* = 1,2,...,5; *j* = 1,2,...,6; *h* = 1,2,...,5.

The matrix $X_{5\times 6}$ and $Y_{5\times 5}$ are transformed into the relative membership matrix of the measured concentration and the relative membership matrix of the standard concentration:

$$R_{5\times6} = \begin{bmatrix} 0.897 & 0.800 & 0.391 & 0.621 & 0 & 0.041 \\ 0.767 & 0.860 & 0 & 0 & 0 & 0.213 \\ 0.875 & 0.617 & 0.250 & 0.808 & 0.458 & 0.383 \\ 0.851 & 0.976 & 0.714 & 0.873 & 0.135 & 0.595 \\ 0.600 & 0.024 & 0.371 & 0.088 & 0 & 0.300 \end{bmatrix} = (r_{ij}),$$

$$S_{5\times5} = \begin{bmatrix} 0 & 0.294 & 0.588 & 0.882 & 1 \\ 0 & 0.333 & 0.667 & 0.833 & 1 \\ 0 & 0.250 & 0.500 & 0.750 & 1 \\ 0 & 0.541 & 0.811 & 0.946 & 1 \\ 0 & 0.294 & 0.588 & 0.882 & 1 \end{bmatrix} = (s_{ih})$$

The matrix $R_{5\times 6}$ is normalized by column to obtain the superscalar weight matrix:

	0.225	0.244	0.227	0.260	0	0.027	
	0.192	0.263	0	0	0	0.139	
$W_{5\times 6} =$	0.219	0.188	0.145	0.338	0.772	0.250	$=(w_{ij})$
	0.213	0.298	0.414	0.365	0.228	0.388	
	0.150	0.007	0.215	0.037	0	0.196	

When the distance parameter p = 2, the calculation procedure is as follows.

Since the value range of k is dynamic, a_j and b_j are not necessarily the same when using formula (5) to calculate different sample j. Taking the sampling point j =1 as an example, the determination of a_j and b_j is briefly described.

When j = 1, from the matrix $R_{5\times6}$, $r_{j=1} = (r_{11}, r_{21}, r_{31}, r_{41}, r_{51})^{T} = (0.897, 0.767, 0.875, 0.851, 0.600)^{T}$ can be obtained. By comparing $r_{11}=0.897$ in vector $r_{j=1}$ with the element value in the first row of matrix $S_{5\times5}$, it can be seen that it is between 0.882 and 1, that is, it falls between levels 4 and level 5. When $r_{21} = 0.767$ is compared with the element value in the second row of matrix $S_{5\times5}$, it is in the interval of 0.667 and 0.833, that is, it falls between level 3 and level 4. Similarly, r_{31} falls between level 4 and level 5, r_{41} falls between level 3 and level 4. In summary, the upper limit value of k is $b_j = 5$, and the lower limit value is $a_j = 3$.

For j = 1, the matrix $W_{5\times 6}$ shows that: $w_{j=1} = (w_{11}, w_{21}, w_{31}, w_{41}, w_{51})^{\mathrm{T}} = (0.225, 0.192, 0.219, 0.213, 0.150)^{\mathrm{T}}$.

Substitute relevant data into formula (5) to obtain: $u_{31} = 0.157$, $u_{41} = 0.604$, $u_{51} = 0.239$.

According to the value conditions of formula (5), $u_{11} = 0$, $u_{21} = 0$.

For j = 2, 3,...,6, the lower limit value a_j and the upper limit value b_j of k are 1 and 5,1 and 3, 1 and 5,1 and 3, 1 and 3 respectively. Thus, when p = 2, the relative membership degree matrix of 6 sampling points for each level is obtained.

$${}_{2}U_{5\times6} = \begin{bmatrix} 0 & 0.005 & 0.030 & 0.016 & 0.102 & 0.024 \\ 0 & 0.016 & 0.531 & 0.055 & 0.375 & 0.869 \\ 0.157 & 0.113 & 0.439 & 0.267 & 0.523 & 0.107 \\ 0.604 & 0.765 & 0 & 0.479 & 0 & 0 \\ 0.239 & 0.101 & 0 & 0.183 & 0 & 0 \end{bmatrix} = \left({}_{2}u_{hj} \right)$$

When the distance parameter p is 1, for j=1, 2,..., 6 similar calculations are made separately, the relative membership matrix of 6 sampling points for each level is obtained

$${}_{1}U_{5\times6} = \begin{bmatrix} 0 & 0.005 & 0.033 & 0.014 & 0.132 & 0.032 \\ 0 & 0.016 & 0.677 & 0.048 & 0.304 & 0.878 \\ 0.206 & 0.108 & 0.290 & 0.426 & 0.564 & 0.090 \\ 0.601 & 0.755 & 0 & 0.380 & 0 & 0 \\ 0.193 & 0.116 & 0 & 0.132 & 0 & 0 \end{bmatrix} = \begin{pmatrix} u_{hj} \end{pmatrix}$$

Thus, the average relative membership degree matrix of p=1 and p=2 sample sets for each level is obtained

	0	0.005	0.031	0.015	0.117	0.028	
	0	0.016	0.604	0.052	0.340	0.873	
$\overline{U}_{5\times 6} =$	0.182	0.110	0.365	0.345	0.543	0.099	$=(\overline{u}_{hj})$
	0.602	0.760	0	0.430	0	0	
	0.216	0.109	0	0.158	0	0	

The level eigenvector of the six sampling points are obtained from formula (6)

 $H = (1,2,3,4,5) \cdot (\overline{u}_{hj}) =$ (4.034, 3.952, 2.334, 3.664, 2.426, 2.071)

According to the level eigenvector H, it can be seen that the soil environmental nutrient levels of the six sampling points are: sampling point 1 level 4; sampling point 2 level 3 to 4; sampling point 3 levels 2 to 3; sampling point 4 levels 3 to 4; sampling point 5 level s2 to 3 and sampling point 6 levels 2 to 3. It can be seen that the soil environmental pollution in this batch of sampling sites is medium or lower, and nutrients have reached a relatively deficient level. Local relevant departments should pay attention to it and take positive measures to improve the soil environment and ensure the sustainable development of agriculture and the economy.

CONCLUSION

The problem of soil environment pollution needs to be solved urgently. Soil environment pollution not only affects the basic quality and production quality of crops, but also causes certain harm to air and water resources, and ultimately harms human life and health through food. Soil environment pollution assessment is of great significance to public health, social security, national food security, and ecological balance.

The fuzzy pattern recognition method of soil environment pollution assessment makes full use of the fuzzy characteristics existing in soil environment assessment, considers the common influence of the weight of evaluation factors and the index value, and also considers the correlation between indexes and the similarity of index characteristics of different sampling points, therefore, the interference of some man-made certain factors is concealed. If conditions permit, this method can add other physical, chemical, and biological indexes, so that the evaluation results can reflect the soil environment conditions more effectively.

The results show that this method is more rigorous in theory, more scientific and accurate in evaluation, and more practical in application, and can provide a basis for ensuring the stability of soil ecological environment, improving the quality of cultivated land, and improving regional ecological conditions. Compared with the membership matrix, the calculation process of this method is more complicated, and still needs to be improved.

In view of the current problems of soil environment pollution, the proposed treatment schemes and measures are as follows: (1) combined with the current situation of agricultural development, it is necessary to further optimize the input structure and production structure of agriculture. (2) Improve relevant laws and regulations, formulate scientific and reasonable prevention and control measures, and strengthen the supervision of soil remediation. (3) Increase the research on soil pollution, actively use advanced soil pollution control technology to improve the soil ecological environment, improve the self-carrying capacity, and self-purification capacity of the ecological environment. (4) Strengthen environmental publicity, and constantly improve the people's awareness of soil environment protection. In the face of environmental protection and economic development, people should pay more attention to the importance of environmental protection. Environmental protection is about safeguarding ourselves and our environment, as well as being responsible for our own safety.

REFERENCES

- Chen, S.Y. 1998. The Fuzzy Sets Theory and Practice for Engineering Hydrology and Water Resources System. Dalian University of Technology Press, Dalian
- Francisca, R. and Cristina, L. 2019. Legal measures to prevent and manage soil contamination and to increase food safety for consumer health: The case of Spain. Environ. Pollut., 250: 883-891.
- Kong, M.Y., Zhong, H.P., Wu, Y.X., Liu, G.D., Xu, Y. and Wang G.X. 2019. Developing and validating intrinsic groundwater vulnerability maps in regions with limited data: A case study from Datong City in China using DRASTIC and Nemerow pollution indices. Environ. Earth Sci., 78(8): 1-14.
- Lv, K.Y. 2018. Spatial and Temporal Variation of Soil Nutrients and Its Influencing Factors: A Case Study on Wudi county. Zhejiang University, Hangzhou.
- Mammo, S., Zhang, K., Chimidi, A. and Ibrahim, H. 2019. Soil quality analysis for the sustainability of forest ecosystem: The case of Chilimo-Gaji Forest, West Shewa Zone, Ethiopia. J. Environ. Earth Sci., 9(3): 1-9.
- Newman, G.S., Coble A.A., Haskins, K.E., Kowler, A.L. and Hart, S.C. 2020.The expanding role of deep roots during long-term terrestrial ecosystem development. J. Ecol., 108(6): 2256-2269.
- Onwosi, C. O., Odimba, J.N., Igbokwe, V.C., Nduka, F.O. and Nwagu, T.N. 2020.Principal component analysis reveals microbial biomass carbon as an effective bioindicator of the health status of petroleum-polluted agricultural soil. Environ. Technol., 41(24): 3178-3190.

Poomagal, S., Sujatha, R., Kumar P.S. and Nvo, D.V. 2020. A fuzzy cognitive map approach to predict the hazardous effects of malathion on the environment (air, water, and soil). Chemosphere, 263: 127926-127926.

Ren, Y., Deng, L.Y., Zuo S.D., Song, X.D. and Li Z.W. 2016. Quantifying the

influences of various ecological factors on the land surface temperature of urban forests. Environmental pollution, 216: 519-529.

Su, H.Z., Yang, M. and Wen, Z.P. 2016. An approach using the multi-factor combination to evaluate high rocky slope safety. Nat. Hazards Earth Syst. Sci., 16(6): 1449-1463.