Applying PPE Model Based on RAGA in Evaluating the Soil Quantity Variation

Fu Qiang[1] , *Xie Yonggang*[2] and *Wang Zhiliang*[1]

[1] Si Chuan University, Chengdu City, China, 610065;
[2] Hei long-jiang University, Harbin City, 150080, China
E-mail: binbin_777@163.net or fuqiang@mail.neau.edu.cn

Abstract: The author apply a new technique of falling dimension named projection pursuit to soil study, through using the improved real coding based accelerating genetic algorithm to optimize the projection direction. Thus, it can transfer multi-dimension data into one dimension data, through searching for the optimum projection direction to realize the soil classification and its grade evaluation. So, the method can avoid the artificial disturb, and acquire preferably effect. Thus, the paper provides a new method to the research of soil classification and grade evaluation.

Keywords: RAGA, PPE, soil quantity, and evaluation

1 Introduction

Soil quality can reflect synthetically the soil characteristic. It is the most sensitive index to post the dynamic movement of soil condition. And soil quality can incarnate the influence caused by human function. (Hu, *et al*., 1999; Zhao, *et al*., 1997) Soil erosion is very severity currently in China. More and more soil resource has been destroyed by human being. Soil quality has degenerated seriously and isn't fitted for cultivation in some area. So, through studying the movement of soil quality under the influence of human being, we not only can provide a theoretical method to describing the change trend of soil quality, but also can post the influence direction of soil environment caused by human. Furthermore, it can prevent soil erosion, improve soil, using limited resource, and advance the agriculture and society to develop continued.

The study of the dynamic movement of soil quality is based on soil quality evaluation, which can reflect the soil quality through the space-time movement of evaluating index. Because there are different objects and index system, so, the evaluating methods are different. The main methods to evaluate the soil quality dynamic movement are multi-variable index KELIGE method and soil quality dynamics method. (Zhao, *et al*., 1997) In China, profess Wang xiaoju applied the method of comparatively soil quality index to study the soil quality movement under artificial cultivation (Wang, *et al*., 1997)

The essential of soil quality evaluation is a synthetically evaluation about different soil sample books according to different index. Because there are several evaluation indexes, so, the evaluation belongs to deal with high-dimension data. The traditional method adopts weight judgement, which is to give a weight to each index to making fuzzy synthetically judgement. During the course of this method, how to choosing the weight is difficult. Thereby, the different weight can induce different result. Recently, an effective technique to falling dimension named projection pursuit (PP) has been raised. The method can put the dot in the high-dimension data space into low-dimension sub-space through projection. Because there are many projection directions during the course of mapping from high-dimension space to low-dimension space, so, the projection direction, which can make the projection index function reaching the maximal value, is the best projection direction. Because there are many indexes that means these indexes should been optimized at the same time. It will be very difficulty to optimizing these indexes at the same time if we adopt traditional method. Now, another new optimization method named genetic algorithm (GA) has risen. The GA method is fit for the problem about multi-dimension and can convergence overall

situation. (Jin, *et al*., 2000) The author puts forward a new method named real coding based accelerating genetic algorithm. (RAGA) Through combining the RAGA with PPE (Projection Pursuit Evaluation) model, the author can use RAGA to optimize the parameters in PPE model at the same time. The result can convert several indexes into one synthetically index. Then, according to the projection value, the stylebooks can be compositor and identified. Thus, we can realize the evaluation of soil quality. The author provides a new method for soil quality evaluation.

2   Projection pursuit evaluation model (PPE)

2.1   Brief introduction of PP model

The main characteristics of PP model are as follows. Firstly, PP model can handle the difficulty named dimension disaster, which have been brought by high-dimension data. Secondly, PP model can eliminate the jamming, which are irrespective with data structure. Thirdly, PP model provides a new approach to handle high-dimension problem using one dimension statistics method. Fourthly, PP method can deal with non-linearity problem.(Jin, *et al*., 2000; Xiang, *et al*., 2000; Zhang. 2000)

2.2   Step of PPE modeling (Zhang. 2000)

The step of building up PPE model includes 4 steps as follows.

Step 1: Normalizing the evaluation indexes set of each stylebook. Now, we suppose the stylebook set is $\{x^*(i,j)|i=1\sim n, j=1\sim p\}$. $x^*(i,j)$ is the index value of j and stylebook of i. n ———the number of stylebook. p ———the number of index. In order to eliminate the dimension influence and unite the change scope of each index value, we can adopt the following formulas to normalize the data.

$$x(i,j) = \frac{x^*(i,j) - x_{\min}(j)}{x_{\max}(j) - x_{\min}(j)} \tag{1a}$$

or:
$$x(i,j) = \frac{x_{\max}(j) - x^*(i,j)}{x_{\max}(j) - x_{\min}(j)} \tag{1b}$$

In formula: $x_{\max}(j)$ and $x_{\min}(j)$ stand for the max and the min of j index value. $x(i,j)$ is the index list after moralization.

Step 2: Constructing the projection index function $Q(a)$. PP method is to turn p dimension data ($\{x^*(i,j)|j=1\sim p\}$) into one dimension projection value $z(i)$ based on projection direction a.

$$a = \{a(1), a(2), a(3), \cdots, a(p)\}, \quad z(i) = \sum_{j=1}^{p} a(j)x(i,j) \qquad (i=1\sim n) \tag{2}$$

Then, we can classify the stylebook according to one-dimension scatter figure of $z(i)$. In formula (2), a stand for unit length vector. Thus, the projection index function can be expressed as follows.

$$Q(a) = S_z D_z \tag{3}$$

In formula: S_z ———the standard deviation of $z(i)$, D_z ———the partial density of $z(i)$.

$$S_z = \sqrt{\frac{\sum_{i=1}^{n}(z(i) - E(z))^2}{n-1}} \tag{4}$$

$$D_z = \sum_{i=1}^{n} \sum_{j=1}^{n} (R - r(i,j)) \cdot u(R - r(i,j))$$ (5)

In formula (4) and (5), $E(z)$ ——the average value of series $\{z(i) \mid i = 1 \sim n\}$; R ——the window radius of partial density, commonly, $R = 0.1 S_z$. $r(i,j)$ ——the distance of stylebook, $r(i,j) = |z(i) - z(j)|$; $u(t)$ ——a unit jump function, if $t \geq 0$, $u(t) = 1$, if $t < 0$, $u(t) = 0$.

Step 3: Optimizing the projection index function. When every indexes value of each stylebook have been fixed, the projection function $Q(a)$ change only according to projection direction a. Different projection direction reflects different data structure characteristic. The best projection direction is the most likely to discovery some characteristic structure of high-dimension data. So, we can calculate the max of $Q(a)$ to estimate the best project direction.

Function: $$\text{Max} : Q(a) = S_z \cdot D_z$$ (6)

Restricted condition: $$\text{s.t.} : \sum_{j=1}^{p} a^2(j) = 1$$ (7)

Formula (6) and (7) is a complex non-linearity optimization, which take $\{a(j) \mid j = 1 \sim p\}$ as optimized variable. Traditional method is very difficulty to calculate. Now, we adopt RAGA to handle the kind of problem.

Step 4: Classification and evaluation. We can put the best projection direction a^* into formula (2), then we can obtain the projection value of each stylebook dot. Compare $z^*(i)$ with $z^*(j)$. If $z^*(i)$ is closer to $z^*(j)$, that means stylebook i and j are trend to the same species. If we dispose $z^*(i)$ from big to small, we can obtain the new stylebook list from good to bad.

3   Real coding based accelerating genetic algorithm (RAGA) (Jin, *et al.*, 2000; Zhou, *et al.*, 2000)

Genetic Algorithm has been put forward by Professor Holland in USA. The main operation includes selection, crossover and mutation. The coding mode of traditional GA adopted binary system. But binary system coding mode has many abuses. So, through consulting literature [6], the author put forward a new method named RAGA. (Real coding based Accelerating Genetic Algorithm) RAGA includes 8 steps as follows. For example, we want to calculate the following best optimization problem.

$$\text{Max} : f(X) \qquad \text{s.t.} : a_j \leq x_j \leq b_j$$

Step 1: In the scope of $[a_j, b_j]$, We can create N group uniformity distributing random variable. **Step 2**: Calculate the target function value and dispose the chromosome from big to small. **Step 3**: Calculate the evaluation function based on order expresses as $eval(V)$. **Step 4:** Selecting operation. **Step 5**: Crossover operation. **Step 6**: Mutation operation. **Step 7**: Evolution iteration. **Step 8**: The above seven steps make up of Standard Genetic Arithmetic. (SGA). But SGA can't assure the whole astringency. Enlightening by reference [6], we can adopt the interval of excellence individual during the course of the first and the second iteration as the new interval. Then, the arithmetic comes into step 1, and runs SGA over again to form accelerate running. Thus, the interval of excellence individual will gradually reduce, and the distance is closer to the best dot. The arithmetic will not stop until the function value of best individual less than a certain value or exceed the destined accelerate times. At this time, the currently group will be destined for the result of RAGA. The above 8 steps make up of RAGA.

4 Application example

4.1 Select the data index and the evaluation grade

Based on the above method, the author adopts time antitheses to analyzing the section of the main cultivated soil in Sanjiang Plain. The author calculates the index projection value of every kinds of topsoil, and evaluates the soil movement trend after reclamation through soil quality synthetically evaluation value.

The paper borrows the data of literature[1], and selects the physical and chemical index. The classification standard of each evaluation index sees also to Table 1. (Hu, *et al*., 1999)

Table 1　Grade system and evaluation index of soil quality in Sanjiang Plain (Topsoil)

Evaluation index	I	II	III	IV	V
organic matter (g/kg)	>80	60～80	40～60	20～40	<20
total nitrogen (g/kg)	>5	3.5～5	2～3.5	0.5～2	<0.5
total phosphor (g/kg)	>2	1.5～2	1～1.5	0.5～1	<0.5
total kalium (g/kg)	>25	17.5～25	10～17.5	2.5～10	<2.5
quick result nitrogen (mg/kg)	>350	275~350	200～275	125～200	<125
quick result phosphor (mg/kg)	>100	70～100	40～70	10～40	<10
quick result kalium (mg/kg)	>350	270～350	190～270	110～190	<110
CEC (coml(+)/kg)	>200	150～200	100～150	50～100	<50
pH	6～6.5	5.5～6, 6.5～7.0	5～5.5	7.0～7.5	<5,<7
porosity (%)	>70	60～70	50～60	40～50	<40
granule content (%)	>60	50～60	40～50	30～40	<30

In the paper, we will evaluate four kinds of soil of four areas. The original data sees also to Table 2. (Hu, *et al*., 1999)

Table 2　The original data of evaluation index

site	Reclamation time	organic matter	total nitrogen	total phosphor	total kalium	quick result nitrogen	quick result phosphor	quick result kalium	CEC	pH	porosity	granule content
		(g/kg)				(mg/kg)		coml(+)/kg			（%）	（%）
Junchuan	Wasteland	27.02	0.66	0.38	3.25	134.00	10.93	76.24	45.75	6.37		
Farm	30 years	4.39	0.28	0.19	2.76	28.00	28.82	52.10	24.24	6.15		
852 Farm	Wasteland	64.32	3.33	1.48	12.38	413.60	6.45	178.21	153.53	6.18		
	30years	31.72	2.01	0.49	0.65	386.40	12.04	103.19	98.20	6.24		
Hulin County	Wasteland	98.97	6.05	1.27	5.86	848.40	12.23	218.80	92.80	5.42		
Crescent	5 years	54.38	3.35	0.91	6.95	355.60	10.50	280.95	86.65	5.36		
nature protect	15 years	34.05	2.82	1.05	8.08	294.00	7.32	115.25	97.83	5.95		
area	25 years	23.24	1.57	0.92	6.75	315.00	9.97	66.16	86.86	5.57		
Baoqing	Wasteland	72.90	5.10	2.20	29.10	209.50	17.00	537.50	169.01	6.21	63.16	78.93
County	2 years	81.40	5.10	2.20	25.10	273.00	91.00	309.00	151.67	6.12	62.33	36.75
Dongsheng	4 years	73.60	4.21	2.40	23.80	287.00	134.80	387.00	136.17	6.35	55.66	29.21
Village	10 years	55.60	3.43	1.61	27.30	251.00	75.40	250.00	10.46	6.17	55.51	34.21

4.2 Build up the PPE model of grade standard

In Table 1, there are 5 grades of soil quality that means 5 stylebooks. Each stylebook has 11 evaluation indexes. The problem belongs to 11 dimension data. Build up PPE model according to Table 1. During the course of RAGA, the parent generation scale is 400. ($n = 400$) The crossover probability is 0.80. ($p_c = 0.80$) The mutation probability is 0.80. ($p_m = 0.80$) The number of excellence individual is 20. $\alpha = 0.05$. Through accelerating 8 times, we can obtain the best projection value. That is 0.4978. The best projection direction: $a^* =$(0.3133, 0.3826, 0.3203, 0.3855, 0.2732, 0.3842, 0.2926, 0.3198, 0.1232, 0.1823, 0.2137). Putting a^* into formula (2), we can obtain the projection value of each soil quality standard stylebook. That are $z_1^*(j) =$(3.1907, 2.7991, 2.0143, 1.2333, 0.8428). About that stylebook which has 9 evaluation indexes, we also adopt the above method. At last, the best projection value is 0.4578. The best projection direction are $a^* =$(0.3322, 0.3974, 0.3323, 0.3997, 0.2914, 0.3975, 0.3028, 0.3335, 0.1253). Putting a^* into formula (2), we can obtain the projection value of each soil quality standard stylebook. That are $z_2^*(j) =$(2.9121, 2.5363, 1.7828, 1.0332, 0.6584).

4.3 Build up PPE model of evaluation stylebook for each area

In Table 2, there are 2 soil stylebooks to be evaluated. Those are before opening up wasteland and the following 30 years. Each stylebook has 9 indexes. Build up PPE model. The parameters in RAGA are the same as before. The soil stylebook projection value are $z_1^*(j) = $ (2.5586, 1.0706). If we compare $z_1^*(j)$ with $z_2^*(j)$, we can know that the soil quality before opening up wasteland belongs to grade II. And the soil quality of the following 30 years belongs to grade IV. That means the soil quality descends seriously. Using the same method, we can build up the PPE model of each farm. The evaluating results see also to Table 3 and Fig.1 to Fig.4.

Table 3   The soil quality evaluation table based on RAGA—PPE model in Sanjiang Plain

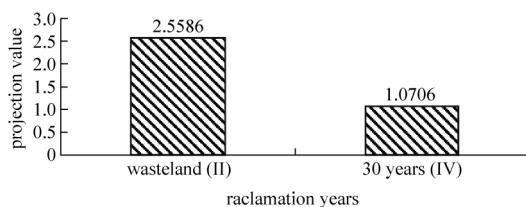|	Junchuan Farm		852 Farm			
	wasteland	30 years	wasteland	30 years	wasteland	5 years
Projection value	2.5586	1.0706	2.4345	1.0168	2.2623	1.5376
Evaluation grade	II	IV	II	IV	II	III
	Hulin County, crescent nature protects area.			Baoqing County, Dongsheng Villege		
	15 years	25 years	wasteland	2 years	4 years	10 years
Projection value	1.0838	0.8818	2.6485	2.6495	2.4495	1.5297
Evaluation grade	IV	IV～V	II	II	III	IV

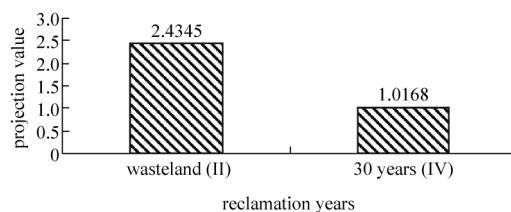**Fig. 1**   Evaluation figure of soil quality movement. (Junchuan Farm)

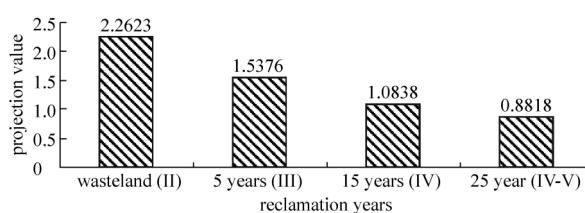**Fig. 2**   Evaluation figure of soil quality movement. (852 Farm)

Fig. 3 Evaluation figure of soil quality movement
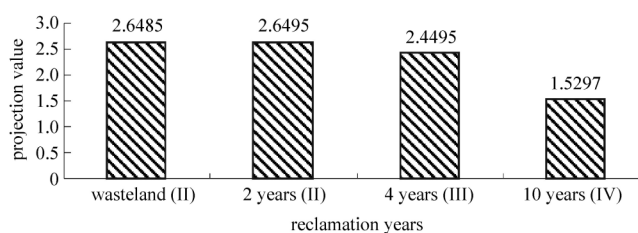(Hulin County, Crescent natural protects area.)

Fig. 4 Evaluation figure of soil quality movement.
(Baoqing county, Dongsheng villege)

5 Analyze the movement tendency of soil quality

Through calculating, we know that the soil quality take on downtrend in general after opened up wasteland in Sanjiang Plain. The result means the soil quality has changed much because of opening up wasteland.

There are different reasons about soil quality degradation in different area. In Junchuan Farm, the main kind of soil is dark brown soil. Wind erosion and water erosion are very serious after reclamation. More and more organic matter and nutrient have been eroded and loosed with soil and water. Furthermore, extensive cultivation for many years and the soil only are used without maintains. So, the soil quality deteriorates step by step. The results make for obviously degenerated. In 852 Farm, the main kind of soil is white slurry soil. The kind of soil has serious wind erosion. The other kinds of soil have been influenced by reasonless cultivation for many years, such as plough and grow in wet soil, soil only can be used without maintain. So, the soil quality degenerate steps by step also.

The main king of soil of Hulin county crescent natural protection area is meadow soil. Meadow soil is the most important cultivated soil. The main reason to degenerate of this king of soil is reasonless cultivated system. Because of the function of drainage after reclamation, the soil moisture reduces, and the humus in topsoil decomposes quickly. Then, a part of nutrient content will be advanced. But many years later, it still takes on degenerated trend as a whole. And that, meadow has its particular landform and water condition, and it is easy to became salinized soil.

The main soil in Baoqing county, Dongsheng village is meadow bog soil. Bog soil shaped in condition of low-lying hypsography and hydrocele perennial. It has the same degeneration mechanism as meadow soil. But bog soil has little soil moisture. The organic matter has been decomposed quickly. The organic matter and other nutrient content are advanced. Despite the soil quality degenerated after reclamation, such as the capability is increasing and the porosity and granule content descend, but the total quality ascends. The phenomena can see also Figure 5(no Fig. 5). The projection value during the course of reclamation original phase is higher than wasteland. But, many years later, because the soil has been used without protection, however, the soil quality will degenerate also.

6 Conclusion

(1) Through applying PPE model, the author builds up the PPE model of soil quality evaluation in

Sanjiang Plain. Several evaluation indexes of soil quality have been taken as multi-dimension projection parameters to seeking the best projection direction. The best projection index function value can reflect the quality of each soil stylebook good or bad. Thus, we can avoid the disturbance by artificial factor to endow weight. The result is good.

(2) The author improves on SGA, and put forward a new method named RAGA. Through reducing the interval of excellence individual to accomplish the accelerate process. Thus, the method of RAGA can realize quick convergence and seeking the best result in the whole scope.

(3) Combing RAGA with PPE model, through using RAGA to optimizing the many parameters in the PPE model, we can obtain the best projection direction of evaluation index of each soil quality stylebook. Thus, the process of PPE modeling has been predigested. And the PPE model can be used in many other fields.

(4) The PPE model has the same good result as literature [1], which adopted soil quality index method. These two methods all can reflect the serious downtrend of soil quality after land reclamation in big scale in Sanjiang Plain. The result indicates the main reason of soil degeneration is soil erosion and reasonless cultivate system. For different kinds of soil and different landform condition, the above two reasons have different function. Thereby, we should grasp the main factor to analyzing synthetically when we study the soil degeneration. Furthermore, the author provides a new method and thought for readers.

References

[1] Hu Jinming, Liu Xingtu, 1999. Evaluate and analyze the soil quality movement in Sanjiang Plain [J]. Geography Science: vol. **19**, No.5 417-421.

[2] Zhao Qiguo, Sun Bo, Zhao Taolin, 1997. Soil quality and persistent environment Ⅰ [M]. Definition and evaluation method of soil quality.[J]. Soil: vol. **29**, No.3 113-120.

[3] Liu Xingtu, 1997. Wetland resource in Songnen Plain and its continuous development [J]. Geography Science: vol. **29**, No. 5 451-460.

[4] Wang Xiaoju, Gong Zitong, Zhang Ximiao, 1997. Workout the varied figure of soil quality based on geography information system [J]. Soil: vol. **29**, No.1 37-42.

[5] Wang Xiaoju, Gong Zitong, 1997. Evaluate and analyze the soil quality movement of different period of time in red soil hill small area [J]. Geography Science: vol. **17**, No.2 141-148.

[6] Jin Juliang, Ding Jing, 2000. Genetic arithmetic and its application to water science [M]. Chengdu: Si Chuan University publishing company: 42-47.

[7] Xiang Jingtian, Shi Jiuen, 2000. The statistics method of data processing in the non-linearity system [M]. Beijing: Science publishing company: 190-192.

[8] Zhang Xinli, 2000. Projection pursuit ant its application to water resources [D]. Chengdu: 67-73.

[9] Zhou Ming, Sun Shudong, 2000. The theory of genetic arithmetic and its application [M]. Beijing: National defense industry publishing company: 4-7,37-38.