

12-7-2011

Regional partitioning of agricultural non-point source pollution in China using a projection pursuit cluster model

XinHu LI

1 State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China;

ChengYi ZHAO

1 State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China; zcy@ms.xjb.ac.cn

Bin WANG

2 College of Water Conservancy and Building Engineering, Northeast Agricultural University, Harbin 100500, China;

Garry FENG

3 Department of Biological Systems Engineering, Washington State University, Washington 99164-6120, USA

Follow this and additional works at: <https://egijournals.researchcommons.org/journal-of-arid-land>



Part of the [Agriculture Commons](#)

Recommended Citation

LI, XinHu; ZHAO, ChengYi; WANG, Bin; and FENG, Garry (2011) "Regional partitioning of agricultural non-point source pollution in China using a projection pursuit cluster model," *Journal of Arid Land*: Vol. 3 : Iss. 4 , Article 6.

DOI: 10.3724/SP.J.1227.2011.00278

Available at: <https://egijournals.researchcommons.org/journal-of-arid-land/vol3/iss4/6>

This Research Article is brought to you for free and open access by Journals of EGI. It has been accepted for inclusion in *Journal of Arid Land* by an authorized editor of Journals of EGI. For more information, please contact hyzhang@ms.xjb.ac.cn.

Regional partitioning of agricultural non-point source pollution in China using a projection pursuit cluster model

Cover Page Footnote

This project was supported by the National Natural Science Foundation of China (40830640), and the Plan for Innovation of Graduate Students of Jiangsu province (CX09B_168Z). We would like to thank the staff of the College of Water Conservancy and Building Engineering for their assistance and cooperation during this study.



Science Press



Regional partitioning of agricultural non-point source pollution in China using a projection pursuit cluster model

XinHu LI¹, ChengYi ZHAO^{1*}, Bin WANG², Garry FENG³

¹ State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China;

² College of Water Conservancy and Building Engineering, Northeast Agricultural University, Harbin 100500, China;

³ Department of Biological Systems Engineering, Washington State University, Washington 99164-6120, USA

Abstract: A projection pursuit cluster (PPC) model was used to analyze the regional partitioning of agricultural non-point source pollution in China. The environmental factors impacting the agricultural non-point source pollution were compiled into a projection index to set up the projection index function. A novel optimization algorithm called Free search (FS) was introduced to optimize the projection direction of the PPC model. By making the appropriate improvements as we explored the use of the algorithm, it became simpler, and developed better exploration abilities. Thus, the multi-factor problem was converted into a single-factor cluster, according to the projection, which successfully avoided subjective disturbance and produced objective results. The cluster results of the PPC model mirror the actual regional partitioning of the agricultural non-point source pollution in China, indicating that the PPC model is a powerful tool in multi-factor cluster analysis, and could be a new method for the regional partitioning of agricultural non-point source pollution.

Keywords: nonpoint pollution; regional partitioning; projection pursuit; Free search

The consumption of chemical fertilizers in China has increased every year since 1978, and China is now the largest consumer and producer of chemical fertilizers in the world (Zhu *et al.*, 2006). In developed countries, if water pollution is to be prevented, the maximum fertilizer N application rate is 150–225 kg N/hm² per season. With an average application rate of 434.3 kg N/hm² per season, China has been the world's largest user of pesticides for more than 10 years, and thus, faces serious environmental issues. The mean pesticide application rate was 8.2 kg/hm² in 2001, which was much higher than the mean rate of developed countries (Zhu *et al.*, 2006). The heavy use of fertilizers and pesticides for intensive farming and low-efficiency cropping systems are often responsible for fertilizer and pesticide leaching into groundwater and runoff (Duan *et al.*, 2000; Xing and Zhu, 2002; Babiker *et al.*, 2003; Liu *et al.*, 2005).

The runoff of fertilizer N and soil N from paddy fields in South China has been found to range from 6.75 to 27.0 kg N/(hm²·a) (Xing and Zhu, 2000). Livestock produc-

tion has developed very rapidly in China, resulting in the total livestock waste in 2002 being over four times greater than production of industrial organic pollutants. Moreover, the burning of straw in the open air causes additional smoke and dust pollution (Xu, 2007).

These pollutants have adverse effects on the environmental and public health, such as eutrophication in lakes and streams, and the pollution of soil and food (Shen, 2005; Zhu *et al.*, 2006). At present, agricultural non-point source pollution has become the dominant source of water pollution in China, and an important factor in air pollution.

Facing these serious environmental problems, researchers have made many suggestions to control the non-point source pollution in China (Zhu, 1992; Widawsky *et al.*, 1998). However, China covers a vast area, and the degree of overuse of fertilizer and pesticide

Received 2011-05-04; accepted 2011-06-28

*Corresponding author: ChengYi ZHAO (E-mail: zcy@ms.xjb.ac.cn)

varies regionally. Remediation should differ in each area, so the regional partitioning of agricultural non-point source pollution in China is very necessary.

Regional partitioning of agricultural non-point source pollution is a typical multi-factor problem. There are many factors that impact agricultural non-point source pollution, such as the application rates of fertilizers and pesticides, soil erosion, and rainfall quantity, intensity and duration.

Analyzing high-dimensional data can be very difficult because (1) it is impossible to draw a visible scatter plot or otherwise graph the data, which would be very helpful for exploratory purposes, and (2) data points are extremely sparse in the high-dimensional space, so kernel smoothers and similar techniques do not work. Projection pursuit (PP) is able to avoid this “curse of dimensionality” since it actually works with a low-dimensional projection. Also, quite often, some irrelevant variables exist when the dimension of a data set is moderate to high (Li and Cheng, 1993).

PP methods look for interesting low-dimensional projection of high-dimensional, multivariate data. The idea originated in a paper by Kruskal (1969), and Friedman and Tukey (1974) were the first to use the term “projection pursuit”. Huber (1985) gave an excellent example of PP that, for the first time, put the fascinating problems and ramifications of PP into a coherent perspective. Cheng and Li (1986) provided an in-depth assessment of the methodology.

PP actually started as a multivariate classification, pioneered by Kruskal (1969), Switzer and Wright (1971), Friedman and Tukey (1974). To date, it has been applied in many fields, such as hydrology (Wang and Ni, 2008), meteorology (Chang *et al.*, 1990), environmental research (Wang, *et al.*, 2004; Wang *et al.*, 2006), and engineering (Van, 1998; Kotas, 2006). These studies concluded that PP is a powerful tool for statisticians to detect hidden non-linear structures in high-dimensional data sets.

The well known Principal components analysis (PCA) belongs to the family of techniques for reducing dimensions. However, PCA only takes into account the second order relationships and may miss interesting relationships that are hidden (Berro *et al.*, 2010). Classical PCA often fails to give reasonable results (Chang *et al.*, 1990; Li and Cheng, 1993). For all that, Caussinus and Ruiz-Gazen (2009) noticed that PP is not widely adopted, when compared with the extensive use of PCA. In the present paper, we take advantage of modern, powerful computers and revisit the field of PP.

The objective of this research is to analyze the regional partitioning of agricultural non-point source pollu-

tion in China using a projection pursuit cluster model. In order to reach our goal, an important condition has to be fulfilled; constructing an optimization algorithm. It is commonly accepted that the role played by the optimization algorithm in the field of PP is crucial. Friedman (1987) pointed out the efficiency of PP strongly relies on the ability of the optimization algorithm to find substantive optima of the projection index among a forest of dummy optima, caused by sampling fluctuations. Therefore, seeking the right projection direction is one of the key issues of the PPC model.

To our knowledge lots of published literature (Guo *et al.*, 2000; Jin and Ding, 2000; Fu *et al.*, 2003; Wang *et al.*, 2004; Berro *et al.*, 2010) used a genetic algorithm (GA) in the context of an exploratory projection pursuit. Only a few papers used the Particle swarm optimization (PSO) and other algorithms. But GA and PSO concepts cannot describe the operation of the algorithm. In particular, they cannot explain free, uncertain, individual behavior (Penev and Littlefair, 2005). A novel optimization algorithm, called Free search (FS), was used in this study to optimize the projection of the PPC model. Some research shows that the optimization ability of the FS algorithm is better than other algorithms, such as GA and PSO (Penev and Littlefair, 2005; Wang and Ni, 2008; Zhou *et al.*, 2009). However, there is no scientific publication using the FS algorithm in the context of exploratory projection pursuit.

1 Materials and methods

1.1 Projection pursuit cluster model

Projection pursuit, a statistical method, was developed to analyze high-dimensional data based on projection (Friedman and Tukey, 1974). High-dimensional data are projected onto one-dimensional space and their characteristics are analyzed through the one-dimensional space.

If $x(i, j)$ ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$, n is the number of samples; m is the number of factors of the samples) is the initial value of the j^{th} factor of the i^{th} sample. The procedures for developing the model are described as follows.

1.1.1 Data standardization

In order to eliminate the effect of different ranges of cluster factor values, the initial data are standardized before they are used in the PPC model. The standardized formula used in this study is:

$$x(i, j) = \frac{x(i, j) - x_{\min}(j)}{x_{\max}(j) - x_{\min}(j)}. \quad (1)$$

Where, $x_{\max}(j)$, $x_{\min}(j)$ are the maximum and minimum

of the j^{th} factor, respectively.

1.1.2 Constructing the projection index function

Essentially, the projection is used to observe data characteristics from all angles. The projection pursuit method is used to integrate m -dimensional data into a one-dimension projection value:

$$z(i) = \sum_{j=1}^m a(j)x(i, j) (i=1, 2, \dots, n). \quad (2)$$

Next, a classification is made in accordance to the one-dimensional dispersion pattern of $\{z(i)|i=1, 2, \dots, n\}$.

An integration index value requires the dispersion of the projection value $z(i)$ to have the partial projection dots cluster as densely as possible. It is even better if the partial dots are condensed into certain dot groups. On the whole, the distance between dot groups should be as wide as possible. Therefore, the projection index function can be expressed as follows:

$$Q(a) = S_z D_z, \quad (3)$$

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

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

Where, $E(z)$ is the average value of the series $\{z(i)|i=1, 2, \dots, n\}$; R is the window radius of partial density, which can be tested, but is normally assumed to equal $0.1 S_z$; and $r(i, j) = |z(i) - z(j)|$ is the distance between the samples, $u(t) = 1$ if $t \geq 0$ and $u(t) = 0$ if $t < 0$.

1.1.3 Model optimization

Different projection directions reflect different data structure characteristics. The best projection direction is the one that may reveal a certain characteristic structure of high-dimensional data. From Eqs. (6) and (7), it is shown that the PPC model reflects an optimum problem. As Friedman pointed out, the efficiency of PP strongly relies on the ability of the optimization algorithm to find substantive optima of the projection index among a forest of dummy optima, caused by sampling fluctuations (Friedman, 1987).

$$\max Q(a) = S_z D_z, \quad (6)$$

$$\sum_{j=1}^m a^2(j) = 1. \quad (7)$$

Free search (FS) was proposed by Penev and Littlefair (2005) as a novel, population-based, optimization method. Free search can model the behavior of animals.

Each animal can operate either with small precise steps for a local search, or with large steps for global exploration. Each animal decides how to search, with small or with large steps. FS can be applied to real value numerical optimization problems, as well as the Genetic algorithm (Holland, 1973; Holland, 1992), the Differential evolution (DE), (Price and Storn, 1997) and the Particle swarm optimization (Kennedy and Eberhart, 1995). In the FS algorithm, the animals have the freedom to decide where and how to search. Free search has similarities with ant colony optimization, modified for continuous search, and space real-coded GA (Jin, 2000), PSO and DE. However, Free Search is not an ant, a swarm or genetic algorithm. The PSO, GA and DE concepts cannot describe the operation of the algorithm.

In particular, they cannot explain a free, uncertain, individual behavior (Penev and Littlefair, 2005), and the research has shown that the FS algorithm is better than other algorithms such as GA, PSO and DE (Penev and Littlefair, 2005; Wang *et al.*, 2008).

1.1.3.1 Free search algorithm architecture

(1) Initialization

$$x_{0ji} = X_{\min i} + (X_{\max i} - X_{\min i}) \text{random}_{ji}(0,1), \quad (8)$$

$$x_{0ji} = a_{ji}, a_{ij} \in [X_{\min i}, X_{\max i}], \quad (9)$$

$$x_{0ji} = c_i, c_i \in [X_{\min i}, X_{\max i}]. \quad (10)$$

Where, x_{0ji} is an initialization location; $\text{random}_{ji}(0,1)$ is a random value between 0 and 1, a_{ji} and c_i are constants; $X_{\min i}$ and $X_{\max i}$ are the search space borders; $j=1, 2, \dots, m$, m is the population size; $i=1, 2, \dots, n$, n is the number of dimensions.

(2) The exploration walk generates coordinates of a new location,

$$x_{tji} = x_{0ji} - \Delta x_{tji} + 2\Delta x_{tji} T \text{random}_{tji}(0,1). \quad (11)$$

Where, x_{tji} is a new location; T is the step limit per walk; t is the current step, and

$t = 1, 2, \dots, T$ $\text{random}_{tji}(0,1)$ is a random value between 0 and 1.

Δx_{tji} is the modification strategy:

$$\Delta x_{tji} = R_{ji}(X_{\max i} - X_{\min i}) \text{random}_{tji}(0,1). \quad (12)$$

Where, R_{ji} is the variable neighbor space $R_{ji} \in [R_{\min}, R_{\max}]$.

The individual behavior, during the walk, is modeled and described as:

$$f_{tj} = f(x_{tji}), f_j = \max(f_{tj}). \quad (13)$$

The pheromone generation is:

$$PH_j = f_j / \max(f_j). \tag{14}$$

The sensibility generation is:

$$SE_j = SE_{\min} + \Delta SE_j. \tag{15}$$

$$\Delta SE_j = (SE_{\max} - SE_{\min}) \text{random}_j(0,1). \tag{16}$$

Where, $SE_{\min} = PH_{\min}$, $SE_{\max} = PH_{\max}$;

The selection and decision-making for a start location, x'_{0ji} for an exploration walk is:

$$x'_{0ji} = x_{ij} (PH_k \geq SE_j), \tag{17}$$

$$k=1, 2, \dots, m; j=1, 2, \dots, m,$$

(3) Acceptable criteria for termination is as follows:

Reaching the optimization criteria:

$$f_{\max} \geq f_{\text{opt}}. \tag{18}$$

$$\text{Expiration of generation limit: } g \geq G. \tag{19}$$

$$\text{Complex criterion: } (f_{\max} \geq f_{\text{opt}}) \parallel (g \geq G). \tag{20}$$

Where, f_{\max} is the maximum achieved solution; f_{opt} is an acceptable value of the objective function; G is a limit and g is a current value.

1.1.3.2 Free Search algorithm improvement

R_{ji} plays an important role in the convergence speeds of the algorithm, but has little effect on the update of an individual location. In addition, a new location is selected according to the relationship between sensibility and pheromone, and the sensibility generation with randomness. The selected exploration walk of a new location is not always the best location. Therefore, the final results of algorithm output are not always optimal.

The FS algorithm was improved by ascertaining the formula (21) (decrease of line) to R_{ji} , where the same R_{it} is adopted per iteration. Thus, the algorithm becomes simpler, and has better exploration abilities.

$$R_{it} = R_{\max} - it(R_{\max} - R_{\min}) / it_{\text{sum}}. \tag{21}$$

Where, it is current iteration times; it_{sum} is the total of the iteration times, $it = 1, 2, \dots, it_{\text{sum}}$.

1.2 Classification

We put the best projection direction a^* into formula (2) to obtain the projection value of the dots of each sample. When comparing $z^*(i)$ with $z^*(j)$, the closer the $z^*(i)$ to $z^*(j)$, the more the samples i and j tend to be in the same class.

2 Results and discussion

Agricultural non-point source pollution is created when rain falls to the earth or snow melts, and the water runs across fields or pavements, taking with it fertilizers, pesticides, waste from livestock, and many other toxic or harmful materials. Soil erosion from poor cropping techniques, runoff of excessive fertilizer or pesticide use, and equipment fuel spills can all result in runoff pollution into nearby streams and lakes, as well as seepage into valuable underground aquifers. According to the analysis above, the index system includes seven factors: (1) Consumption of chemical fertilizer (kg/hm²) (CF); (2) Consumption of pesticide (kg/hm²) (P); (3) Production of wastes from livestock (kg/hm²) (L); (4) Production of crop straw (kg/hm²) (CS); (5) Rainfall amount (mm) (R); (6) Ratio of hydraulic erosion area to land area (HL), and (7) Ratio of agricultural acreage to land area (AL) (Table 1). The factors' statistical values are shown in Table 1 (Editorial Department of China Agricultural Yearbooks, 1991–2007).

The projection pursuit was used to conduct regional partitioning of agricultural non-point source pollution for 30 administrative regions in China, listed in Table 1. Let the initial species scale 'm' equal 20, and the step limit per walk 't' equal 20. The correct projection direction, on the basis of the projection pursuit, is as follows: $a^* = (0.3605 \ 0.5171 \ 0.1914 \ 0.2541 \ 0.4455 \ 0.2867 \ 0.4700)$. The projection value was obtained by adding a^* into formula (2). The scatter diagram of the projection values is shown in Fig. 1, providing obvious proof that 5 regions can be partitioned.

The cluster results are shown in Table 2, listed as followed:

(1) Region 1 (Shandong province): Shandong province is known as the worst and most severe area for agricultural non-point source pollution in China. Shandong is one of the most well-known vegetable-producing regions in China, and also hosts the largest distribution center for vegetables grown in north China, especially in Shouguang city. In 2001, the vegetable production area was about 1.96×10^4 hm². However, excessive nutrient application is typical in this region. For example, the average fertilizer N application rate is over 1,000 kg N/hm² (Zhu and Chen, 2002), which is much higher than

Table 1 Data points of agricultural non-point pollution

Administrative regions	Index						
	CF	P	L	CS	R	HL	AL
Beijing	465	14.68	7,113	7,409	561.6	26.75	0.02011
Tianjin	467	6.71	12,633	7,170	569.3	3.98	0.04421
Hebei	345	9.19	11,625	6,246	524.9	29.10	0.04672
Shanxi	252	6.01	3,424	4,836	504.7	59.31	0.02442
Inner Mongolia	188	2.38	1,345	5,701	265.7	13.14	0.00551
Liaoning	316	12.05	6,731	8,580	804.0	32.97	0.02575
Jilin	279	5.83	6,340	9,016	595.8	10.1	0.02608
Heilongjiang	150	4.71	2,591	5,361	531.6	19.12	0.00231
Shanghai	357	20.76	7,763	9,517	1,089.4	0.00	0.05009
Jiangsu	446	13.51	5,641	8,661	986.4	3.97	0.07358
Zhejiang	332	23.13	2,791	8,089	1,601.3	17.75	0.02759
Anhui	311	10.34	8,228	6,096	1,172.1	13.39	0.06525
Fujian	492	22.59	3,282	7,443	1,645.5	12.11	0.01988
Jiangxi	246	14.34	4,496	7,658	1,638.0	21.03	0.02682
Shandong	436	14.59	17,372	8,755	795.0	20.64	0.09029
Henan	372	7.55	20,925	7,509	747.5	18.16	0.04438
Hubei	393	15.13	6,046	8,317	1,173.1	37.72	0.04342
Hunan	263	14.2	8,328	8,304	1,447.0	19.07	0.02284
Guangdong	425	18.07	6,308	7,509	1,771.0	6.14	0.03600
Guangxi	310	8.21	6,826	6,381	1,532.7	4.38	0.00347
Hainan	479	23.26	9,130	5,416	1,749.6	0.60	0.02404
Sichuan	233	5.94	5,045	7,337	973.8	31.09	0.01998
Guizhou	161	2.03	8,934	5,622	1,112.7	41.55	0.02757
Yunnan	236	5.05	4,464	5,342	1,265.8	37.21	0.01604
Xizang	179	3.09	102	7,882	588.0	5.22	0.00019
Shaanxi	351	2.35	3,487	4,793	780.8	57.40	0.02065
Gansu	204	5.58	2,285	4,852	301.1	29.50	0.00925
Qinghai	147	3.66	1,216	5,697	277.4	7.41	0.00068
Ningxia	272	1.45	3,518	5,796	260.6	40.37	0.02177
Xinjiang	289	3.9	79	8,808	144.9	7.04	0.00233

Table 2 Cluster results of 30 administrative regions by projection pursuit

No.	Administrative regions	Projection value	Cluster	Region	No.	Administrative regions	Projection value	Cluster	Region
1	Shandong	-2.7665	1	1	16	Tianjin	-0.3057	3	
2	Fujian	-1.8575	2	2	17	Guangxi	0.50592	4	
3	Guangdong	-1.7686	2		18	Sichuan	0.51548	4	
4	Hubei	-1.7553	2		19	Shaanxi	0.58446	4	
5	Jiangsu	-1.7383	2		20	Jilin	0.59685	4	4
6	Shanghai	-1.7376	2		21	Yunnan	0.66452	4	
7	Hainan	-1.7190	2		22	Guizhou	0.73389	4	
8	Zhejiang	-1.6678	2		23	Shanxi	0.77406	4	
9	Henan	-0.85573	3	3	24	Ningxia	1.49320	5	
10	Anhui	-0.83534	3		25	Xinjiang	1.97120	5	
11	Hunan	-0.79422	3		26	Gansu	2.05830	5	
12	Jiangxi	-0.76754	3		27	Xizang	2.25830	5	5
13	Beijing	-0.60616	3		28	Heilongjiang	2.33720	5	
14	Liaoning	-0.53258	3		29	Inner Mongolia	2.65020	5	
15	Hebei	-0.33194	3		30	Qinghai	2.89620	5	

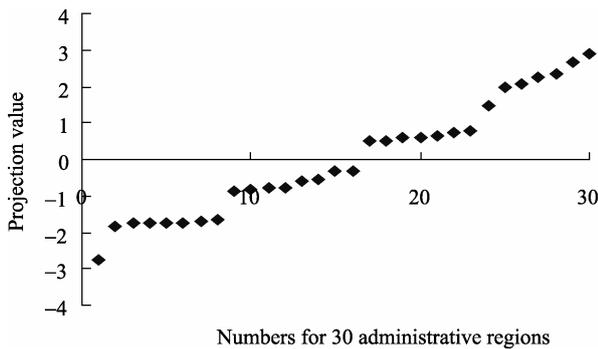


Fig. 1 Scatter diagram of projection value

crop demands. Consequently, nitrate pollution in the ground water has been monitored in this intensive vegetable-producing area.

Nitrogen accumulation commonly becomes serious in vegetable production systems. The average N in the 0–90 cm soil profile in greenhouses in Shouguang was 709 kg N/hm² (Zhu *et al.*, 2006). Moreover, Shandong is a major grain producing area, where an increase in grain production has resulted from greater use of chemical fertilizers. Therefore, Shandong is the most severely polluted (from agricultural non-point sources) area in China.

(2) Region 2 includes seven regions; Zhejiang province, Hainan province, etc. These regions are major fruit and vegetable producing areas, with slightly better conditions than Shandong. The high market value of vegetables and fruit promote the over-fertilization and overuse of pesticides (Ye, 2001). Moreover, these areas are mainly located in eastern China, where abundant rain makes the transference of pollutants by leaching and erosion easier.

(3) Region 3 includes eight regions; Henan province, Anhui province, etc. These regions are major grain producing areas, which continue to use chemical fertilizers and pesticides (Huang and Rozelle, 1995), since there is a significant correlation between the annual fertilizer application and increased grain production.

(4) Region 4 includes seven regions; Guangxi province, Sichuan province, etc. These regions are mainly newly developing areas, whose lower soil productivity and relatively lower agricultural production (Guo, 2005), equates to less use of fertilizers and pesticides.

(5) Region 5 includes seven regions; Ningxia Hui

autonomous region, Qinghai province, etc. These regions have the lowest agricultural non-point source pollution in China. They are located in Western China, where the rainfall is low, causing less runoff. A lack of water resources is the main factor limiting larger-scale agricultural production, especially in Ningxia Hui autonomous region and Gansu province. Effective irrigation systems (such as drip irrigation, spray irrigation, etc.) have been used in these areas, especially in Xinjiang Uygur autonomous region (Lei *et al.*, 2005). Therefore, leaching seldom occurs. Because these areas have the lowest fertilizer and pesticide over usage rates, their degree of agricultural non-point source pollution is the lowest in China.

3 Conclusion

This study used a projection pursuit cluster model to analyze the regional partitioning of agricultural non-point source pollution in China. The results show that five regions have been partitioned, and Shandong province was the worst area for agricultural non-point source pollution.

In conclusion, the results of the projection pursuit mirrors the real world agricultural non-point source pollution data in China, indicating that the projection pursuit cluster model is a powerful tool in multi-factor problem and could be a new method for the regional partitioning of agricultural non-point source pollution in China.

The Free search algorithm (FS) is a novel, population-based optimization method, which was used to optimize the projection direction. FS can be more reliable for the optimization of unknown problems because it can find the best projection direction more accurately than other methods.

Acknowledgement

This project was supported by the National Natural Science Foundation of China (40830640), and the Plan for Innovation of Graduate Students of Jiangsu province (CX09B_168Z). We would like to thank the staff of the College of Water Conservancy and Building Engineering for their assistance and cooperation during this study.

References

Babiker I S, Mohamed A A, Tetao H, *et al.* 2003. Assessment of groundwater contamination by nitrate leaching from intensive vegetable cultivation using geographical information system.

Environment International, 29(8): 1009–1017.

Berro A, Larabi Marie-Sainte S, Ruiz-Gazen A. 2010. Genetic algorithms and particle swarm optimization for exploratory projection

- pursuit. *Annals of Mathematics and Artificial Intelligence*, 60(1–2): 153–178.
- Caussinus H, Ruiz-Gazen A. 2009. *Exploratory Projection Pursuit. Data Analysis (Digital Signal and Image Processing Series)*. London: ISTE, 85–111.
- Chang H, Shi J, Chen Z. 1990. Projection pursuit principal component analysis and its application to meteorology. *Acta Meteorol Sinica*, 4: 254–263
- Cheng P, Li G. 1986. Projection pursuit—a kind of new statistical method. *Chinese Journal of Applied Probability and Statistics*, 2: 267–276.
- Duan S W, Zhang S, Huang H Y. 2000. Transport of dissolved inorganic nitrogen from the major rivers to estuaries in China. *Nutrient Cycling in Agroecosystems*, 57(1): 13–22.
- Editorial Department of China Agricultural Yearbooks. 1991–2007. *China Agriculture Yearbook*. Beijing: China Agriculture Press.
- Friedman J H, Tukey J W. 1974. A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, C-23: 881–890.
- Friedman J H. 1987. Exploratory projection pursuit. *Journal of the American Statistical Association*, 82: 249–266.
- Fu Q, Xie Y G, Wei Z M. 2003. Application of projection pursuit evaluation model based on real-coded accelerating genetic algorithm in evaluating wetland soil quality variations in the Sanjiang plain, China. *Pedosphere*, 13(3): 249–256.
- Guo Q, Questier F, Massart D L, *et al.* 2000. Sequential projection pursuit using genetic algorithms for data mining of analytical data. *Analytical Chemistry*, 72(13): 2846–2855.
- Guo X M. 2005. Sichuan province agriculture: main limited factors and suggestions. *Rural Economy*, (5): 3–4.
- Holland J H. 1971. Genetic algorithms and the optimal allocations of trials. *Siam Journal on Computing*, 2(2): 88–105.
- Holland J H. 1992. Genetic algorithms. *Scientific American*, 4: 44–50.
- Huang J K, Rozelle S. 1995. Environment stress and grain yields in China. *American Journal of Agricultural Economics*, 77: 853–864.
- Huber P J. 1985. Projection pursuit. *Annals of Statistics*, 13(2): 435–475.
- Jin J L, Ding J. 2000. *Research on Applications of Genetic Algorithm to Water Resources Engineering*. Msc Thesis: Sichuan University, 28–31
- Kennedy J, Eberhart R C. 1995. Particle swarm optimization. In: *IEEE International Conference on Neural Networks, Perth of Australia*. Piscataway: IEEE Service Centre, 1942–1948.
- Kotas M. 2006. Application of projection pursuit based robust principal component analysis to ECG enhancement. *Biomedical Signal Process Control*, 1(4): 289–298.
- Kruskal J B. 1969. *Toward A Practical Method Which Helps Uncover The Structure Of A Set Of Multivariate Observations By Finding The Linear Transformation Which Optimizes A New Index Of Condensation*. New York: Computational Statistics, 427–440.
- Lei X Y, He C M, Wei X J, *et al.* 2005. Research on operating mechanisms of the water-saving agriculture in Xinjiang. *Journal of Irrigation and Drainage*, 24(2): 44–47.
- Li G Y, Cheng P. 1993. Some recent developments in projection pursuit in China. *Statistic Sinica*, 3: 35–51.
- Liu G D, Wu W L, Zhang J. 2005. Regional differentiation of non-point source pollution of agriculture-derived nitrate nitrogen in groundwater in northern China. *Agriculture, Ecosystems & Environment*, 107: 211–220.
- Penev K, Littlefair G. 2005. Free search—a comparative analysis. *Information Science*, 172: 173–193.
- Price K, Storn R. 1997. Differential evolution—a simple evolution strategy for fast optimization. *Dr Dobb's Journal*, 22(4): 18–24.
- Shen R P, Sun B, Zhao Q G. 2005. Spatial and temporal variability of N, P and K balances in agroecosystems in China. *Pedosphere*, 15(3): 347–335.
- Switzer P, Wright R M. 1971. Numerical classification applied to certain Jamaican Eocene nummulitids. *Mathematical Geology*, 3: 297–331.
- Van Hulle M M. 1998. Non parametric regression modeling with equiprobable topographic maps and projection pursuit learning with application to PET image processing. *Journal of VLSI Signal Processing Systems for Signal Image and Video Technology*, 18 (3): 275–285.
- Wang B, Zhang Z Y, Zhang G H, *et al.* 2008. Free search: a novel algorithm for optimizing irrigation scheduling. *Advance Water Science*, 19(5): 736–741.
- Wang S J, Yang Z F, Ding J. 2004. Projection pursuit cluster model and its application in water quality assessment. *Journal of Environment Science*, 16(6): 994–995.
- Wang S J, Ni C J. 2008. Application of projection pursuit dynamic cluster model in regional partitioning of water resources in China. *Water Resource Management*, 22: 1421–1429.
- Wang S J, Zhang X L, Yang Z F, *et al.* 2006. Projection pursuit cluster model based on genetic algorithm and its application in Karstic water pollution evaluation. *Journal of Environment Pollution*, 29(3–4): 256–260.
- Widawsky D, Rozelle S, Jin S Q. 1998. Pesticide productivity host-plant resistance and productivity in China. *Agricultural Economics*, 19: 203–217.
- Xing G X, Zhu Z L. 2000. An assessment of N loss from agriculture fields to the environment in China. *Nutrient Cycling Agroecosystems*, 57(1): 67–73.
- Xing G X, Zhu Z L. 2002. Regional nitrogen budgets for China and its major watersheds. *Biogeochemistry*, 57–58(1): 405–427.
- Xu Y H. 2007. Straw incineration pollution and countermeasures in China. *Environment and Sustainable Development*, (3): 21–23.
- Ye Q M. 2001. Chemical strategies to reduce pesticide contamination in citrus cultivation in Zhejiang. *Pesticide Science and Administration*, (Suppl.): 39–41.
- Zhou H, Shao S H, Xu C. 2009. Comparison between free search and particle swarm optimization. *Microelectronics & Computer*, 26(5): 9–12.
- Zhu Z L. 1992. Efficient management of nitrogen fertilizers for flooded rice in relation to nitrogen transformations in flooded soils. *Pedosphere*, 2(2): 97–114.
- Zhu Z L, Chen D L. 2002. Nitrogen fertilizer use in China—contribution to food production, impacts on environment strategies. *Nutrient Cycling Agroecosystems*, 63(2): 97–114.
- Zhu Z L, Norse D, Sun B. 2006. *Policy for Reducing Non-point Source Pollution from Crop Production in China*. Beijing: China Environmental Science Press, 65–75.