Journal of Environmental Sciences Vol.15, No.4, pp.464—468, 2003 ISSN 1001—0742 CN11—2629/X

Article ID: 1001-0742(2003)04-0464-05 **CLC number**: X51; X11 **Document code**: A

Optimizing the atmospheric sampling sites using fuzzy mathematic methods

FANG Chun-sheng* , WANG Ju, DONG De-ming, YU Lian-sheng
(College of Environment and Resources, Jilin University, Changchun 130023, China. E-mail: fangcs@mail.jlu.edu.cn)

Abstract: A new approach applying fuzzy mathematic theorems, including the Primary Matrix Element Theorem and the Fisher Classification Method, was established to solve the optimization problem of atmospheric environmental sampling sites. According to its basis, an application in the optimization of sampling sites in the atmospheric environmental monitoring was discussed. The method was proven to be suitable and effective. The results were admitted and applied by the Environmental Protection Bureau (EPB) of many cities of China. A set of computer software of this approach was also completely compiled and used.

Keywords: fuzzy set; primary matrix element theorem; fisher classification method; optimization of sampling sites

Introduction

The higher resolution in both absolute space and time in environmental monitoring is very difficult to be achieved, and is not necessary sometimes either. The leading principle of environmental monitoring is to obtain the typical monitoring data of the largest scope of space by the least number of sampling sites (Chief Environmental Monitoring Station of China, 1992). Thus optimization methods and theorems of how sampling sites are to be selected are becoming increasingly important. The study on the optimization of atmospheric sampling sites in China began in 1980s, especially after the Environmental Monitoring Technological Guideline was conferred in 1987.

In the early of 1980s, Doctor R. E. Munn from Canada made a detailed discussion on the optimization methods of atmospheric environment sampling sites and summarized that there were mainly 4 kinds of optimization methods. They are functional zone based methods, statistics based methods, systematic simulation based methods and the forecasting technology based comprehensive methods.

Almost all of the statistics based optimization methods are based on the grid monitoring data. After many attempts and modifications, we put forward a set of optimization methods of the atmospheric environment sampling sites. The methods including many fuzzy mathematic theorems, such as the primary matrix element theorem and the similarity principle, and many statistic theorems, i.e. Fisher classification method, t-test and F-test methods.

The software, optimal analysis system of sampling sites in the atmosphere environment (Yu, 1994), was developed successfully on the base of the theorems and methods mentioned above using C++ programming language (for Windows). This set of optimization methods and the software taken on and recommended by National EPB of China (Now SEPA). It has been applied in many cities' EPB.

1 Methods

1.1 Primary matrix element theorem

After professor L. A. Zadeh provided the concept of fuzzy set in 1965 (Zimmermann, 1984), he forecasted that fuzzy set could be used in the classification problem. He researched on the classification problem with a fuzzy equivalence matrix. The primary matrix element theorem (PMET) was first submitted in 1982 (Yu, 1992). It is described as follows:

1. If field X is a finite set, a fuzzy matrix $\underset{\sim}{R} = (r_{ij})_{n \times n}$ is defined on field X; 2. set the diagonal element of $\underset{\sim}{R}$ to 1, i.e. $r_{ii} = 1$; 3. $\underset{\sim}{R}$ is a symmetrical matrix, i.e. $r_{ij} = r_{ji}$; A fuzzy relation matrix

* Corresponding author

meeting the above 3 conditions is called a fuzzy similar matrix; 4. $\underset{\sim}{R}$ is a transitive matrix, i.e. $\underset{\sim}{R} = \underset{\sim}{R} \times \underset{\sim}{R}$; A $n \times n$-dimensions fuzzy matrix $\underset{\sim}{R} = (r_{ij})_{n \times n}$ that fulfill the above conditions is called a fuzzy equivalence matrix, $\underset{\sim}{R} \in \Phi(X)$; 5. let $\lambda_i = \bigvee_{\substack{j=1 \\ j \neq i}}^{n} (r_{ij}), \lambda_i$, is a $n \times 1$ vector, it represents the confidence level (CL) of the *ith* row. We assume X is the whole body of the atmosphere environmental sampling sites, where $x_i = \{x_1, x_2, \cdots x_n\}$ is the sampling sites vector and $y_i = \{y_1, y_2, \cdots y_m\}$ is the sampling date vector, so when $X(x, y)$ is a finite set, we can obtain a fuzzy equivalence matrix $\underset{\sim}{R}$, to represent the fuzzy relationship between the ordered pair (x, y). Then we can obtain the confidence level of each sampling sites x_i. The sorted confidence level of each sampling site can be used in the latter classification procedure.

1.2 Fisher classification method

The fisher classification method(Sankar, 1986) is usually applied to resolve the classification problem of a series of the ordered data.

Suppose that the ordered data are $x_1, x_2, \cdots x_n$ (where $x_1 \geqslant x_2 \geqslant \cdots \geqslant x_n$). Assume a possible classification group of each data x_i is $[x_i, x_{i+1}, \cdots x_j]$ $(j \geqslant i)$, the mean value of this group is defined as $\bar{x}_{ij} = \frac{1}{j - i + 1}\sum_{k=i}^{j} x_k$, and the group diameter is defined as $D(i, j) = \sum_{k=i}^{j}(x_k - \bar{x}_{ij})^2$. The group diameter reflects the variance among the data in the given group.

Suppose we will separate the n ordered data into k groups, one possible classification result is $p(n, k): [i_1 = 1, i_1 + 1, \cdots, i_2 - 1][i_2, i_2 + 1, \cdots, i_3 - 1]\cdots[i_k, i_k + 1, \cdots, n]$, the variable i_k in the classification result is the subscript of each data, $i_1 = 1 < i_2 < \cdots < i_k \leqslant n$. The error function of this classification result is defined as $e[p(n, k)] = \sum_{j=1}^{k} D(i_j, i_{j+1} - 1)$.

When n and k is fixed, the less the value of $e[p(n, k)]$, the better is the classification result. The Fisher classification is a method that can cause the error function, $e[p(n, k)]$, to reach its least value and obtain the corresponding classification result of the given ordered data.

1.3 The computing software

In order to optimize the sampling sites in the atmosphere environment of a certain city in the middle of China, we designed the air pollutant sampling schemes. 48 sampling sites were set at 1 km × 1 km scale per site. Because total suspended particle(TSP) is the typical pollutant of the city, from March 18 to April 10, 1998 we carried out the atmosphere monitoring for TSP (Table 1).

2 Materials and methods

Data treatment procedure

Generally, there are seven steps in the calculating procedure:

First, standardize the source sampling data matrix $C = (c_{ij})_{n \times m}$ in Table 1, and we get matrix $A = (a_{ij})_{n \times m}$, where $a_{ij} = \frac{c_{ij}}{c_s}$.

Where c_s is the 2nd level of environmental air quality standard of TSP (i.e. 0.3 mg/Nm^3 in GB3095-96).

Second, obtain the fuzzy similar matrix $\underset{\sim}{R} = (r_{ij})_{n \times n}$.

There are two sub-steps:

(1) Set the non-diagonal elements by affiliated function: $r_{ij} = \sum_{k=1}^{m} \alpha \times a_{ik} \times a_{jk}$.

Where $\alpha = \max\limits_{\substack{i=1\\ j=1\\ i\neq j}}^{n}\left(\sum_{k=1}^{m} a_{ik} \times a_{jk}\right)$ is a constant to ensure r_{ij} in the range of $[0, 1]$; a_{ik}, a_{jk} are the standardization values calculated in the first step; thus, $r_{ij} \in [0,1]$, $r_{ij} = r_{ji}$.

Table 1 The sampling data of each sampling sites in the certain city (mg/Nm³)

	3/18	3/19	3/20	3/21	3/24	3/25	3/26	3/29	3/30	3/31	4/1	4/2	4/3	4/4	4/5	4/6	4/7	4/8	4/9	4/10
1 #	0.281	0.354	0.426	0.488	0.283	0.141	0.131	0.375	0.392	0.630	0.547	0.625	0.390	0.108	0.255	0.272	0.365	0.164	0.486	0.447
2 #	0.268	0.328	0.233	0.485	0.341	0.172	0.220	0.256	0.271	0.270	0.260	0.349	0.442	0.112	0.195	0.260	0.548	0.287	0.528	0.360
3 #	0.298	0.328	0.228	0.262	0.327	0.229	0.298	0.384	0.480	0.293	0.306	0.335	0.837	0.291	0.208	0.580	0.538	0.442	0.350	0.769
4 #	0.150	0.125	0.154	0.079	0.125	0.115	0.097	0.167	0.228	0.166	0.231	0.238	0.208	0.238	0.172	0.210	0.211	0.262	0.158	0.109
5 #	0.243	0.261	0.179	0.204	0.453	0.228	0.141	0.330	0.404	0.446	0.135	0.275	0.474	0.072	0.116	0.253	0.385	0.209	0.257	0.178
6 #	0.205	0.114	0.061	0.096	0.237	0.198	0.140	0.242	0.182	0.306	0.285	0.480	1.069	0.594	0.259	0.480	0.282	0.764	0.229	0.926
7 #	0.102	0.190	0.245	0.200	0.230	0.101	0.111	0.182	0.259	0.330	0.255	0.433	0.338	0.080	0.178	0.286	0.230	0.150	0.321	0.209
8 #	0.197	0.339	0.808	0.321	0.258	0.159	0.680	0.326	0.425	0.494	0.518	0.320	0.424	0.212	0.138	0.204	0.458	0.286	0.443	0.421
9 #	0.258	0.293	0.394	0.344	0.476	0.241	0.197	0.447	0.657	0.743	0.498	0.505	0.745	0.142	0.315	0.440	0.548	0.287	0.528	0.360
10 #	0.120	0.167	0.089	0.154	0.258	0.274	0.227	0.244	0.194	0.143	0.270	0.278	0.437	0.087	0.180	0.244	0.382	0.208	0.278	0.243
11 #	0.173	0.182	0.246	0.167	0.307	0.156	0.058	0.258	0.275	0.358	0.348	0.359	0.491	0.093	0.173	0.237	0.177	0.243	0.141	0.279
12 #	0.356	0.300	0.262	0.233	0.436	0.277	0.236	0.250	0.164	0.267	0.229	0.378	0.589	0.212	0.200	0.238	0.464	0.560	0.317	0.315
13 #	0.228	0.242	0.264	0.312	0.397	0.296	0.298	0.393	0.298	0.458	0.496	0.564	0.686	0.140	0.298	0.425	0.626	0.376	0.486	0.492
14 #	0.224	0.174	0.248	0.199	0.261	0.212	0.185	0.224	0.296	0.247	0.256	0.234	0.294	0.047	0.171	0.248	0.429	0.223	0.285	0.298
15 #	0.050	0.068	0.050	0.032	0.160	0.186	0.145	0.174	0.222	0.217	0.214	0.223	0.226	0.089	0.110	0.267	0.251	0.227	0.292	0.414
16 #	0.222	0.157	0.237	0.206	0.200	0.098	0.098	0.194	0.267	0.262	0.430	0.311	0.504	0.044	0.168	0.147	0.306	0.123	0.192	0.214
17 #	0.176	0.185	0.285	0.158	0.169	0.150	0.252	0.295	0.589	0.547	0.443	0.360	0.452	0.160	0.288	0.441	0.605	0.339	0.386	0.294
18 #	0.156	0.228	0.204	0.153	0.176	0.101	0.087	0.301	0.391	0.542	0.382	0.366	0.534	0.052	0.112	0.262	0.317	0.202	0.251	0.255
19 #	0.259	0.280	0.296	0.259	0.250	0.141	0.170	0.249	0.287	0.364	0.394	0.450	0.484	0.112	0.220	0.224	0.308	0.320	0.428	0.274
20 #	0.232	0.213	0.232	0.183	0.298	0.119	0.188	0.237	0.255	0.271	0.397	0.380	0.485	0.112	0.166	0.229	0.414	0.244	0.288	0.229
21 #	0.169	0.161	0.170	0.169	0.202	0.144	0.129	0.207	0.222	0.242	0.362	0.298	0.370	0.071	0.214	0.226	0.352	0.230	0.137	0.180
22 #	0.314	0.251	0.124	0.200	0.490	0.244	0.204	0.222	0.153	0.244	0.340	0.514	0.772	0.112	0.188	0.102	0.269	0.216	0.288	0.279
23 #	0.157	0.180	0.234	0.149	0.247	0.158	0.149	0.130	0.235	0.162	0.277	0.251	0.253	0.092	0.125	0.211	0.296	0.172	0.268	0.198
24 #	0.078	0.132	0.154	0.090	0.186	0.120	0.070	0.132	0.134	0.155	0.149	0.158	0.215	0.048	0.034	0.261	0.218	0.171	0.212	0.338
25 #	0.290	0.250	0.237	0.200	0.320	0.198	0.209	0.381	0.246	0.301	0.414	0.852	0.486	0.171	0.302	0.234	0.294	0.433	0.403	0.372
26 #	0.289	0.412	0.333	0.324	0.281	0.147	0.137	0.364	0.620	0.332	0.362	0.332	0.474	0.124	0.223	0.295	0.439	0.316	0.358	0.451
27 #	0.197	0.372	0.209	0.198	0.189	0.097	0.146	0.110	0.277	0.458	0.497	0.488	0.498	0.084	0.141	0.158	0.267	0.287	0.593	0.108
28 #	0.121	0.185	0.277	0.169	0.196	0.115	0.125	0.200	0.279	0.232	0.238	0.251	0.291	0.088	0.182	0.209	0.238	0.166	0.287	0.272
29 #	0.244	0.268	0.347	0.274	0.347	0.139	0.138	0.362	0.366	0.408	0.458	0.473	0.500	0.106	0.212	0.236	0.362	0.238	0.337	0.384
30 #	0.253	0.241	0.336	0.240	0.336	0.159	0.064	0.266	0.280	0.947	0.357	0.468	0.335	0.167	0.122	0.204	0.290	0.205	0.465	0.328
31 #	0.597	0.456	0.242	0.219	0.331	0.134	0.324	0.267	0.160	0.428	0.452	0.306	0.364	0.131	0.160	0.182	0.458	0.298	0.670	0.386
32 #	0.248	0.280	0.152	0.198	0.300	0.217	0.192	0.294	0.322	0.420	0.331	0.262	0.404	0.102	0.113	0.220	0.546	0.285	0.301	0.358
33 #	0.114	0.099	0.156	0.150	0.341	0.204	0.176	0.167	0.132	0.155	0.355	0.345	0.513	0.137	0.125	0.161	0.390	0.260	0.350	0.237
34 #	0.151	0.110	0.128	0.216	0.229	0.132	0.074	0.238	0.468	0.286	0.466	0.414	0.314	0.161	0.168	0.219	0.267	0.180	0.360	0.208
35 #	0.272	0.179	0.166	0.226	0.283	0.224	0.194	0.295	0.228	0.455	0.397	0.457	0.410	0.189	0.188	0.242	0.524	0.295	0.452	0.392
36 #	0.394	0.475	0.233	0.274	0.590	0.184	0.174	0.435	0.464	0.475	0.394	0.604	1.047	0.246	0.120	0.216	0.542	0.446	0.484	0.684
37 #	0.300	0.386	0.204	0.224	0.398	0.365	0.143	0.328	0.207	0.475	0.293	0.429	0.626	0.175	0.151	0.200	0.443	0.228	0.415	0.390
38 #	0.268	0.286	0.107	0.310	0.187	0.148	0.130	0.204	0.140	0.194	0.268	0.315	0.384	0.118	0.162	0.193	0.218	0.446	0.284	0.297
39 #	0.153	0.486	0.084	0.104	0.280	0.131	0.108	0.156	0.141	0.225	0.202	0.134	0.606	0.040	0.084	0.211	0.399	0.227	0.286	0.433
40 #	0.175	0.316	0.163	0.120	0.245	0.148	0.170	0.236	0.219	0.207	0.272	0.248	0.454	0.176	0.206	0.231	0.382	0.325	0.278	0.221
41 #	0.174	0.169	0.175	0.133	0.258	0.152	0.176	0.237	0.163	0.160	0.273	0.196	0.380	0.104	0.179	0.218	0.308	0.222	0.286	0.200
42 #	0.203	0.193	0.174	0.121	0.256	0.101	0.171	0.369	0.229	0.254	0.335	0.298	0.440	0.490	0.166	0.266	0.360	0.215	0.327	0.320
43 #	0.291	0.160	0.128	0.177	0.254	0.204	0.185	0.187	0.207	0.191	0.233	0.106	0.355	0.104	0.182	0.285	0.308	0.165	0.283	0.312

Table 1 (Cont'd)

	3/18	3/19	3/20	3/21	3/24	3/25	3/26	3/29	3/30	3/31	4/1	4/2	4/3	4/4	4/5	4/6	4/7	4/8	4/9	4/10
44 #	0.191	0.213	0.258	0.266	0.277	0.270	0.210	0.170	0.222	0.281	0.248	0.283	0.299	0.140	0.128	0.230	0.319	0.428	0.514	0.375
45 #	0.160	0.082	0.113	0.145	0.221	0.200	0.080	0.191	0.157	0.222	0.267	0.313	0.283	0.097	0.125	0.263	0.251	0.227	0.222	0.401
46 #	0.178	0.170	0.127	0.114	0.149	0.056	0.066	0.096	0.200	0.140	0.114	0.106	0.412	0.057	0.062	0.294	0.160	0.216	0.254	0.372
47 #	0.045	0.088	0.142	0.087	0.100	0.050	0.020	0.342	0.204	0.276	0.269	0.189	0.228	0.040	0.224	0.260	0.316	0.149	0.296	0.290
48 #	0.108	0.160	0.287	0.241	0.062	0.057	0.072	0.274	0.334	0.328	0.219	0.270	0.250	0.061	0.191	0.464	0.332	0.158	0.293	0.279

Note: the above data are daily average concentrations

(2) Set the diagonal elements by function: $r_{ii} = 1(i = 1, 2, \cdots, n)$.

Thus the fuzzy similar matrix $\underset{\sim}{R}$ is obtained.

Third, use the $t(\underset{\sim}{R}) = \underset{\sim}{R} \times \underset{\sim}{R}$ algorithm to make a fuzzy equivalence matrix $\underset{\sim}{R}$ (the matrix is omitted for the limit of this paper's length).

Fourth, the confidence level vector of the fuzzy equivalence matrix $\underset{\sim}{R}$ that constructed by the above 3 steps can be obtained and used to represent the pollution intensity index of each sampling sites (Table 2).

Fifth, sort the confidence level values of each sampling sites and separate them into 5 groups by their pollution intensity with the Fisher classification method (Table 3 and Fig. 1). We separated the 48 sampling sites into 5 groups according to the relationship between the classification number k and the classification error measured by $e[p(n,k)]$ (Fig. 2).

Table 2 The confidence levels (CL) of the 48 sampling sites

No.	CL	No.	CL	No	CL
1 #	0.8365	17 #	0.7829	33 #	0.5699
2 #	0.7211	18 #	0.6343	34 #	0.7151
3 #	0.9514	19 #	0.6807	35 #	1.0000
4 #	0.3748	20 #	0.6221	36 #	0.7835
5 #	0.6337	21 #	0.4957	37 #	0.5488
6 #	0.9345	22 #	0.7166	38 #	0.5979
7 #	0.5242	23 #	0.4477	39 #	0.5684
8 #	0.8179	24 #	0.3664	40 #	0.4859
9 #	0.9783	25 #	0.7834	41 #	0.6155
10 #	0.5291	26 #	0.7765	42 #	0.4942
11 #	0.5764	27 #	0.6535	43 #	0.6028
12 #	0.7477	28 #	0.4729	44 #	0.4772
13 #	0.9252	29 #	0.7411	45 #	0.4255
14 #	0.5405	30 #	0.7386	46 #	0.4348
15 #	0.4280	31 #	0.7579	47 #	0.5233
16 #	0.5399	32 #	0.6624	48 #	0.5699

Table 3 The Fisher classification result of the 48 sampling sites

No.	CL	Group	No.	CL	Group	No.	CL	Group
36 #	1.0000	1	22 #	0.7166	2	14 #	0.5405	
9 #	0.9783	$\bar{x}$ = 0.9759	35 #	0.7151		16 #	0.5399	
3 #	0.9514	S = 0.0310	19 #	0.6807		10 #	0.5291	
6 #	0.9345		32 #	0.6624		7 #	0.5242	4
13 #	0.9252		27 #	0.6535		48 #	0.5233	
1 #	0.8365		18 #	0.6343	3	21 #	0.4957	
8 #	0.8179		5 #	0.6337	$\bar{x}$ = 0.6337	43 #	0.4942	
37 #	0.7835		20 #	0.6221	S = 0.0277	41 #	0.4859	
25 #	0.7834		42 #	0.6155		45 #	0.4772	
17 #	0.7829	2	44 #	0.6028		28 #	0.4729	
26 #	0.7765	$\bar{x}$ = 0.7630	39 #	0.5979		23 #	0.4477	
31 #	0.7579	S = 0.0381	11 #	0.5764		47 #	0.4348	5
12 #	0.7477		34 #	0.5699	4	15 #	0.4280	$\bar{x}$ = 0.4284
29 #	0.7411		40 #	0.5684	$\bar{x}$ = 0.5356	46 #	0.4255	S = 0.0406
30 #	0.7386		33 #	0.5664	S = 0.0305	4 #	0.3748	
2 #	0.7211		38 #	0.5488		24 #	0.3664	

Sixth, calculate the mean value and the variance of each group (Table 3), pick out all the sampling sites of each group whose confidence level is within the range of $[\bar{x} - S/2,\ x + S/2]$ (Table 4). Where $\bar{x}$ and S are the mean value and standard deviation of sampling site's CLs in each group. Then we can obtain all of the pre-optimizing schemes by the combination of the selected sampling sites of each group. There are 216 pre-optimizing schemes in total.

Finally, conduct a mean value t-test, variance t-test, and other tests between each of the pre-optimizing schemes and the population. Considering the functional region information and other conditions, we obtained the finial optimization of the sampling sites (Table 5).

Table 4 The pre-optimizing schemes

	Group 1	Group 2	Group 3	Group 4	Group 5
Sites	3 #,	26 #, 31 #, 12 #	18 #, 5 #, 20 #	38 #, 14 #, 16 #, 10 #, 7 #, 48 #	23 #, 47 #, 15 #, 4 #
Number	1	3	3	6	4

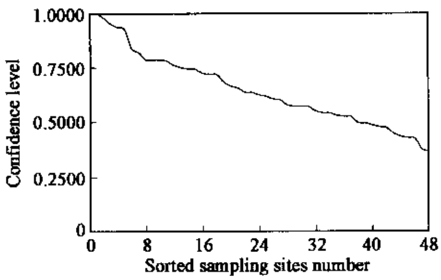

Fig. 1 The confidence level of sampling sites in the city

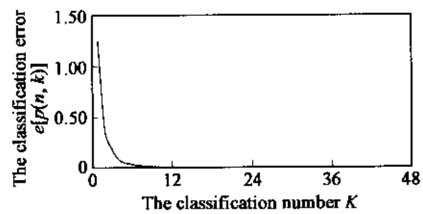

Fig. 2 Relationship between the classification error $e[p(n,k)]$ and the classification number K

Table 5 Final optimizing scheme and statistics tests

Group	Optimizing sampling site	Mean and t-test	Variance and F-test	Population mean and variance
Group 1	3 #			
Group 2	31 #	$\bar{x} = 0.281940$	$S_{Opt} = 0.077006$	$\bar{x} = 0.269192$
Group 3	5 #	$t = 0.395197$	$F = 1.273757$	$S_{Pop} = 0.004634$
Group 4	38 #	$t_{0.05}(48+5-2) = 1.64$	$F_{0.05}(47,4) = 2.53$	
Group 5	23 #			

3 Conclusions

In this paper, we presented a set of theorems and methods to optimize the sampling sites in the atmospheric environmental monitoring. The relative bias between the mean of the optimization scheme and the population is only 4.74% (which meets the 10% limit recommended by the Chief EPB of China), and the t-test and F-test are not statistically significant.

References:

Chief Environmental Monitoring Station of China, 1992. The optimization methods of atmospheric environment sampling sites[M]. Beijing: Environmental Science Press of China.

Liu C M, Xu Z L, Du Y G *et al*., 2000. Analyses of volatile organic compounds concentrations and variation trends in the air of Changchun, the Northeast of China[J]. Atmospheric Environment, 34(26):4459—4466.

Sankar K P, Dwijesh K, Dutta M, 1986. Fuzzy mathematical approach to pattern recognition[M]. Calcutta, India: Indian Statistical Institute.

Yu L S, 1992. Fuzzy system and its application in environment[M]. Changchun: Jilin University Press.

Yu L S, Fang C S, Wang J *et al*., 1996. The application of fuzzy sets approaches to regional noise monitoring[J]. Acta Scientiarum Naturalium Universitatis Jilinensis, 118(4): 71—74.

Yu L S, Li W, Fang C S *et al*., 1994. The optimizing system of the atmospheric environment sampling sites[J]. Environmental Monitoring in China, 10(4): 3—5.

Yu L S, Wang X, Wang J *et al*., 1997. Study on the sustainable developing index system of regional environment[J]. Environmental Science in China, Supplement: 7—10.

Zimmermann H J, 1984. Fuzzy set theory and its applications[M]. USA: Kluwer - Nijhoff Publishing.

(Received for review April 25, 2002. Accepted June 18, 2002)