CHINESE GEOGRAPHICAL SCIENCE
Volume 1, Number 2, pp.154–165, 1991
Science Press, Beijing, China

NUMERICAL CLASSIFICATION OF REGIONAL STRUCTURE OF AGRO–ECOSYSTEM IN GUANGDONG ①

Luo Manjia (罗曼嘉)
(South China Agricultural University, Guangzhou 510642, PRC)

ABSTRACT: On the basis of "The Statistical Data of Guangdong Province in 1981" and the 14 selected original variables, 107 counties and suburbs in Guangdong are classified. In the classification, cluster analysis is applied, and the cluster hierarchy of the agro–ecosystem is achieved with a computer. According to output level, planting structure, and input level of energy, there is an evident regionalism in the agro–ecosystem of Guangdong, which presents approximately concentric circles centralized by Guangzhou City.
KEY WORDS: agro–ecosystem, regional structure, cluster analysis, location quotient

One of the characteristics of an agro–ecosystem is obvious regionalism. Thus , it is necessary to examine their regional structures in the study of macro–system . This paper discusses the regional structure of agro–ecosystem in Guangdong using the method of cluster analysis.

I. METHODS

The essential data of the research is selected from "The Statistical Data of Guangdong Province in 1981". According to the 14 original variables, 107 counties or suburbs in Guangdong as basic units are classified with a microcomputer.

1. Selection of Classifying Variables

In order to reflect the agricultural structures of counties as fully as possible, along with considering output levels, planting structures, and input levels of energy in these

① In the paper, Guangdong includes Guangdong Province and Hainan Province. The latter was established in 1988.

subsystems, the 14 original variables are used as following:

Basic conditions:

(1) farmland area per capita *(FAPC)*

Output level:

(2) grain yield per capita *(CYPC)*

(3) sugar cane yield per capita *(SYPC)*

(4) peanut yield per capita*(PYPC)*

(5) jute yield per capita*(JYPC)*

(6) pig on hand per capita*(PHPC)*

(7) fish yield per capita *(FYPC)*

Planting structure:

(8) area for rice / area for grain*(R / G)*

(9) area for potato / area for grain *(P / G)*

(10) area for sugar cane / farmland area *(S / F)*

(11) area for tea / farmland area *(T / F)*

(12) area for orange / farmland area *(O / F)*

Input level of energy:

(13) machine power / farmland area *(MP / F)*

(14) chemical fertilizer / farmland area *(CF / F)*

In order to ensure the independency of variables, these variables are first sifted on the basis of their correlativity. As a result, correlation coefficients between S / F and $SYPC$, MP / F and $FYPC$, and R / G and P / G exceed ± 0.5 (respectively 0.84, 0.68 and −0.65). Thus, three variables $SYPC$, $FYPC$ and P / G are abandoned because they can be approximately described by S / F, MP / F and R / G. From here, an original matrix is set up and 107 basic units as classified points are located in the space of 11 dimensions, which consists of 11 variables as listed above.

2. Cluster Analysis

Cluster analysis can include three main processes: data standardization, similarity measurement and cluster mergence. Every process can have alternative approaches. In this paper, the respective methods are data normalization—Euclidean distance measurement—sum to error square[1].

2.1. Data normalization

Following formula is used:

$$Xij = (aij - aj\ (min.)) \ / \ (aj\ (max.) - aj\ (min.))$$
$$i = 1,2,\cdots,107$$
$$j = 1,2,\cdots,11$$

Here, aij is an element of the original matrix. After the normalization, a standardized data matrix is obtained, whose dimension is 107×11, i.e. $X_{107 \times 11}$.

2.2 Similarity measurement

The key to cluster analysis is the selection of a statistical variable by which the similarity between any two points can be measured. This variable can be distance coefficients, association coefficients, correlation coefficients, or probabilistic similarity coefficients. In the paper, the Euclidean distance is used to measure the similarity.

$$D_{i,k} = \left(\sum_{j=1}^{11} (X_{ij} - X_{kj})^2 / m\right)^{0.5}$$

$$i,k = 1,2,\ldots,11 \qquad m = 11$$

Here, $D_{i,k}$ is an Euclidean distance between any two points in the space of 11 dimensions. Form the measurement for 107 points, a similarity triangle matrix is obtained as follows:

$$\begin{matrix} D_{1,2} & D_{1,3} & D_{1,4} & \cdots & D_{1,107} \\ & D_{2,3} & D_{2,4} & \cdots & D_{2,107} \\ & & D_{3,4} & \cdots & D_{3,107} \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ & & & & \cdot \\ & & & & D_{106,107} \end{matrix}$$

2.3 Clustering

Initially, 107 points can be considered as 107 clusters (or groups), each of them containing only one element. According to the similarity triangle matrix, a pair of the most similar clusters (p and q) can be found and merged. Thus , the number of clusters is reduced by one. Then , the original standardized matrix is renewed by the following formula:

$$X_{tj} = (N_p X_{pj} + N_q X_{qj}) / (N_p + N_q)$$

Here, t is a new cluster after the mergence, and Np and Nq are respectively the element numbers that p and q contain. This process is repeated until the number of clusters becomes one. Thus all elements are included by a cluster.

When clusters are merged, sum to error square of kth cluster is calculated and signed 1_k. The sum to error square of all clusters is $\sum_{lk}$. For every mergence, the increment of $\sum_{lk}$ must be minimum. The values $\sum_{lk}$ and the order of mergences are recorded one by one so that a cluster hierarchy is obtained.

II. RESULT AND ANALYSIS

Through 106 times of clustering, we get the cluster hierarchy of agro-ecosystem in Guangdong (Fig. 1). From Fig. 1, when the level of similarity measurement is adopted as four, eight clusters can be divided (Table 1) . This paper will mainly analyze and discuss on this level.

Table 1 Cluster attributes, and numbers of counties and suburbs in Guangdong Province

No. of clusters	Element numbers	Names and No. of elements
1	12	Guangzhou(1),Jiangmen(49),Huaxian(2),Conghua(3),Gaohe(45),Nanhai(42),Enping(48), Taishan(46),Kaiping(47),Xinhui(44), Foshan(50),Shaoqing(61)
2	18	Zengcheng(5),Sihui(52),Gaoyao(51), Fengkai(55),Xinxing(60),Meixian(21), Zhongshan(39),Lianxian(67),Wongyuan(73), Qingyuan(72),Deqing(56),Fogang(74), Lechang(66),Yunfu(59),Gaozhou(81), Yangshan(70),Yingde(71),Qujiang(107)
3	4	Panyu(4),Sanshui(43), Doumen(40),Shunde(41)
4	9	Longmen(6),Shixing(65),Renhua(62), Lianshan(68),Huizhou(29),Huiyang(30), Nanxiong(64),Dongguan(32),Boluo(33)
5	21	Xuwen(75),Dingan(93),Chengmai(95), Changjiang(104),Suixi(77),Haikou(88), Wenchang(89),Dunchang(94),Lingao(96), Haikang(76),Qonghai(91),Lingshui(98), Baoting(101),Qongzhong(105),Wanning(92), Aixian(99),Baisha(100),Ledong(102), Qongshan(90),Zhanxian(97),Dongfang(103)
6	23	Xinfeng(7),Pingyuan(26),Huaiji(54), Yunan(57),Luoding(58),Dapu(24), Fengshun(25),Jiaoling(27),Xinyi(82), Jieyang(15),Huilai(20),Xingning(22), Wuhua(23),Longchuan(35),Zijin(36), Heyuan(34),Lianping(37),Liannan(69), Shantou(8),Chaoyang(14),Meizhou(28), Huazhou(79),Wuchuan(80)
7	8	Chaozhou(9),Chaoan(11),Raoping(12), Chenghai(10),Puning(17),Nanao(13), Heping(38),Guangning(53)
8	12	Jiexi(16),Maoming(87),Dianbai(83), Haifeng(18),Lufeng(19),Zhanjiang(86), Ruyuan(63),Lianjiang(78),Huidong(31), Yangjiang(84),Yangchun(85),Shaoguan(106)

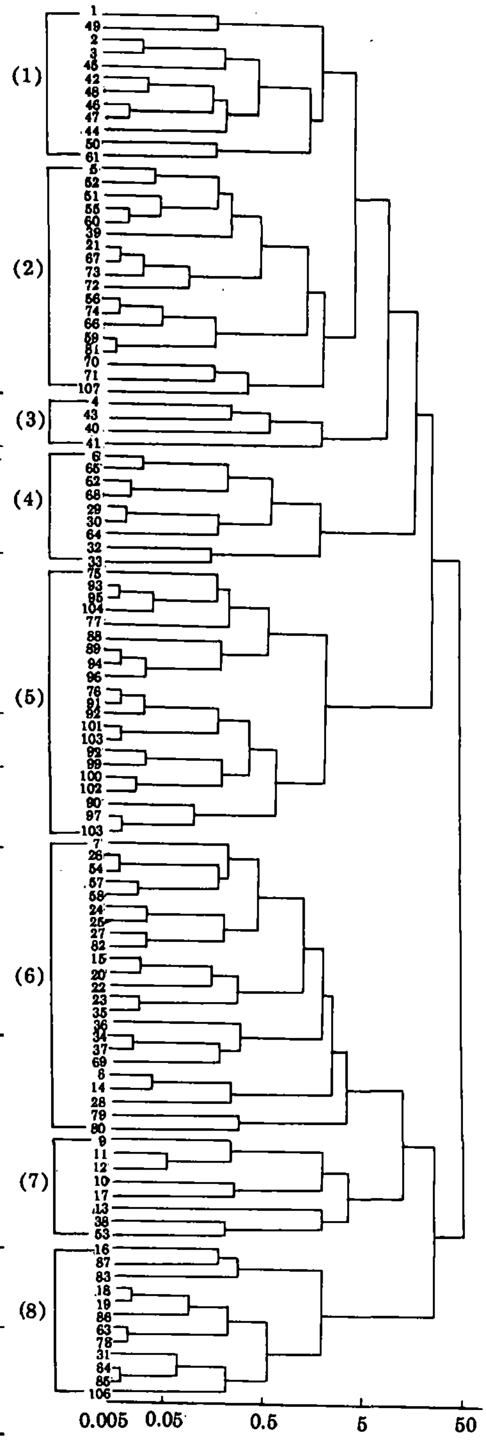

Fig. 1 Cluster hierarchy of agro-ecosystem in Guangdong

1. The Different Characteristics of Eight Clusters

In order to distinguish the differences among these clusters, the averages of 11 variables in respective clusters are calculated using the following formula:

$$\overline{Y}_{jk} = a / N_k \cdot \sum_{p=1}^{N_k} X_{ij}$$

$$j = 1,2,\ldots,11$$

$$i = 1,2,\ldots,8$$

Here, N_k is the number of elements that Kth cluster contains. As a result, "average matrix" $Y_{jk}$is obtained as shown in Table 2. After the average matrix is standardized by the following formula:

$$Z_{jk} = \overline{Y}_{jk} / \sum_{k=1}^{8} \overline{Y}_{ik}$$

the standardized matrix of variables in eight clusters is obtained as shown in Table 3. Based on the standardized matrix, distributing diagrams for 11 variables can be drawn and used to describe the characteristics of eight clusters (Fig.2).

Table 2 Average matrix of varialbes in eight clustres

Varialbes	Clutser							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FAPC(a.)	7.0	7.2	11.3	9.9	10.4	4.7	3.3	6.5
GYPC(kg)	392.5	437.5	500.0	495.5	308.5	286.5	239.5	321.5
PYPC(kg)	14.8	9.5	17.6	22.8	4.7	7.8	13.0	14.5
HYPC(kg)	0.33	0.21	0.74	0.49	0.08	2.00	0.82	1.60
PHPC(H.)	0.59	0.43	0.51	0.37	0.51	0.31	0.37	0.38
R / G(%)	95	86	95	92	75	79	73	74
S / F(%)	5.72	3.54	32.27	3.63	9.12	1.71	2.83	4.87
T / F(%)	0.66	1.82	0.23	0.42	0.08	2.08	4.92	0.64
O / F(%)	1.50	0.89	0.24	1.28	0.06	1.12	3.55	0.41
MP / E(hp / a.)	44	18	33	20	12	17	38	18
CF / (kg / a.)	519	485	593	320	108	395	390	323

From Fig.2 (a), it can be seen that the level of grain output and farmland area are approximatively corresponding with each other. For example, farmland area in cluster 7 is small, and its grain yield per capita is low too. Fig.2 (b) shows the important regions of peanut and jute production. In Fig.2 (c), there are three crest values for pig on hand per capita, and it can be seen that the levels of energy input (CF / F) in cluster 1 and 3 are higher, as well as their pig on hand per capita. This shows a high level of agricultural

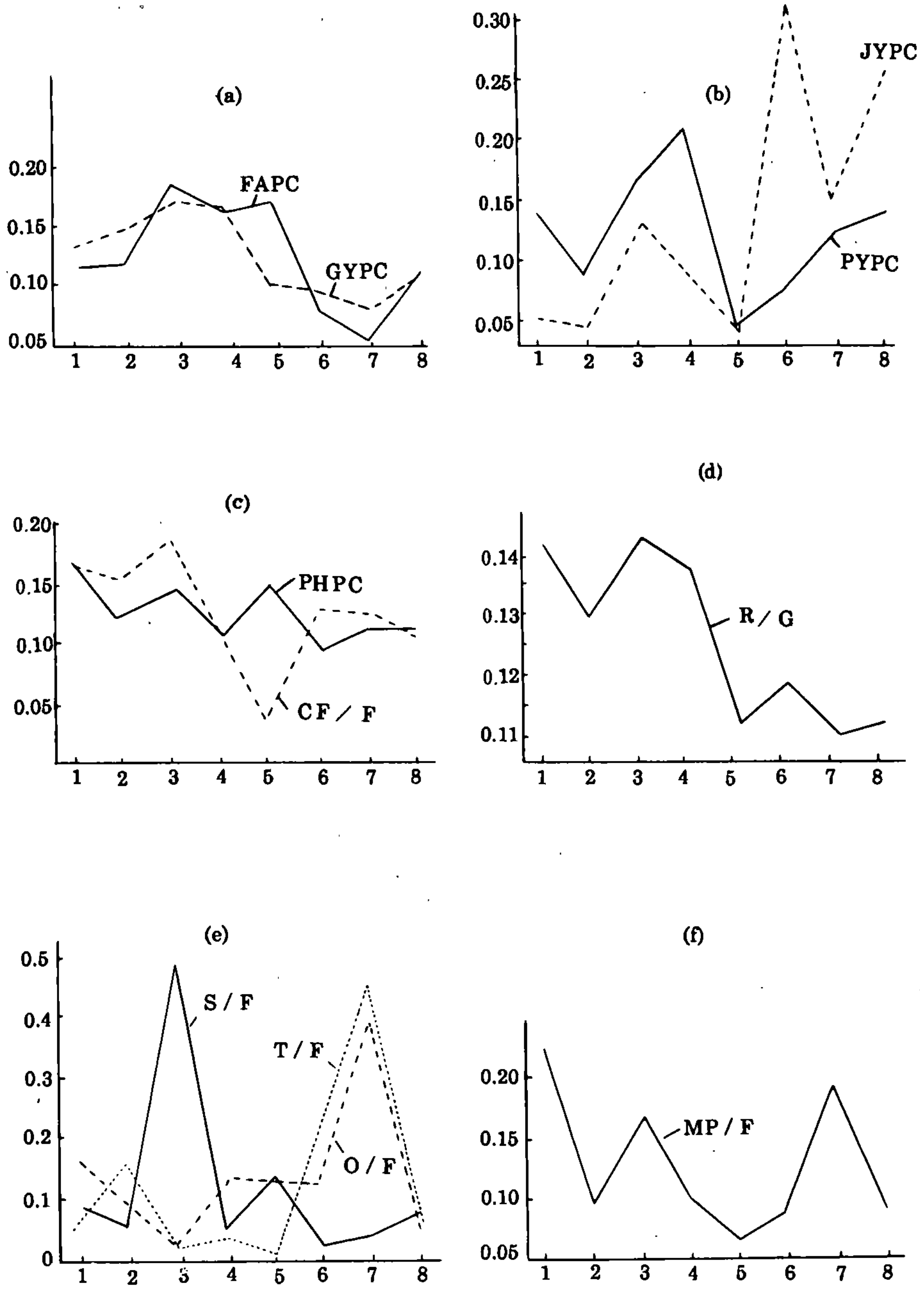

Fig. 2 (a) comparison of farmland and grain production; (b) peanut and jute production; (c) comparison of output and input; (d) rice production; (e) production of sugarcane, tea and orange; and (f) input of machinery.

Table 3 Total standardized values of variables in 8 clusters

Variables	Cluster (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FAPC	0.116	0.119	0.108	0.165	0.173	0.078	0.054	0.107
GYPC	0.132	0.147	0.168	0.166	0.103	0.096	0.080	0.108
PYPC	0.142	0.091	0.168	0.217	0.045	0.074	0.124	0.138
JYPC	0.052	0.034	0.119	0.078	0.012	0.319	0.131	0.255
PHPC	0.169	0.125	0.147	0.105	0.146	0.080	0.108	0.111
R / G	0.142	0.129	0.124	0.137	0.111	0.118	0.109	0.111
S / F	0.090	0.056	0.507	0.057	0.143	0.027	0.044	0.077
T / F	0.061	0.168	0.021	0.039	0.007	0.192	0.454	0.059
O / F	0.166	0.098	0.027	0.141	0.006	0.124	0.393	0.045
MP / F	0.217	0.092	0.164	0.099	0.062	0.086	0.191	0.090
CF / F	0.166	0.155	0.189	0.102	0.024	0.126	0.125	0.103

production—high input and high output—in cluster 1 and 3. In the secondary production based on cultivation, the marketable pig as an output is corresponding with pig on hand. However, cluster 5 , which is regionally in Hainan Island, is characterized by high level of pig on hand and low level of energy input. The reason is that extensive pig-breeding in Hainan Island doesn't depend on the cultivation. Thus , the output of marketable pig is low, even though pig on hand is high there. In 1983, for example, the marketable rate of pig in Shunde (in cluster 3) is 53% , but the rate in Dongfang (in cluster 5) is only 20%. Therefore, the entire level of agricultural production in cluster 5 is low—low input and low output.

From Fig.2 (d) and (e) , it is very easy to see the major or specialized regions of rice, sugar cane, tea and orange production. Cluster 1-4 are major regions of rice production; cluster 5 is an important region of sugar cane production; and cluster 7 is a specialized region of both tea and orange production.

In Fig.2 (f) , there are three crest values of machinery input—cluster 1,3 and 7, which are respectively located in the developed Zhujiang (Pearl) River Delta and the Chaoshan Plain, where superior geographical locations are advantageous to the introduction of auxiliary energy.

2. The Relationship of Different Clusters

From the standardized matrix (Table 3) , it is obvious that:

$$\sum_{k=1}^{8} Z_{jk} = 1$$

Thus, the mean of standardized matrix is:

$$\overline{Z} = (\sum_{k=1}^{8} Z_{jk}) / 8 = 0.125$$

Based on the value $\overline{Z}$, (i.e. $M_{jk}=1$ when $Z_{jk} \geqslant \overline{Z}$ and $M_{jk}=0$ when $Z_{jk} \leqslant \overline{Z}$), a dual matrix of eight clusters can be obtained as Table 4 shows. This dual matrix can describe the general characteristics and differences between any two clusters as well.

Table 4 Dual matrix of 8 clusters

Variables	Cluster							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FAPC	0	0	1	1	1	0	0	0
GYPC	1	1	1	1	0	0	0	0
PYPC	1	0	1	1	0	0	0	1
JYPC	0	0	0	0	0	1	1	1
PHPC	1	1	1	0	1	0	0	0
R / G	1	1	1	1	0	0	0	0
S / F	0	0	1	0	1	0	0	0
T / F	0	1	0	0	0	1	1	0
O / F	1	0	0	1	0	0	1	0
MP / F	1	0	1	0	0	0	1	0
CF / F	1	1	1	0	0	1	1	0

From Fig.1, a classifying structure of eight clusters can be drawn as Fig. 3 shows, Connecting Fig. 3 with Table 4, the relative model of eight clusters can be developed as Table 5 shows.

III. DISCUSSION ON THE REGIONAL STRUCTURE

1. The Characteristics of Spatial Distribution

The regional structure of agro–ecosystem in Guangdong can be mapped as shown in Fig. 4. It is clear that cluster 1 and 3 are located in the developed Zhujiang River Delta and Northwest River Delta; they enjoy the best geographical advantages in the system. Cluster 4 lies in the East River Delta, and the hollow basin in the northern Guangdong; Cluster 6 and 2 are symmetrically located in the east or west of Cluster 4; Cluster 8 is located in the coastal regions centered by the Zhujiang River Delta; Cluster 7 is mainly situated in the Chaoshan Plain; Cluster 5 regionally includes Hainan Island and Leizhou Peninsula.

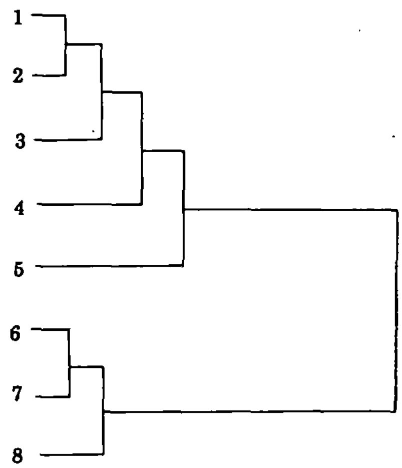

Fig. 3 Classifying structure of 8 clusters

Table 5 The relationship among 8 clusters

<table>
<tr><th rowspan="2">Variables</th><th colspan="8">Cluster</th><th rowspan="2">Variables</th></tr>
<tr><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th></tr>
<tr><td>JYPC</td><td colspan="5"></td><td colspan="3">1</td><td>JYPC</td></tr>
<tr><td>GYPC
R/G</td><td colspan="4">1</td><td>0</td><td colspan="2">1</td><td>0</td><td>T/F
CF/F</td></tr>
<tr><td>PHPC
CF/F</td><td colspan="3">1</td><td>0</td><td></td><td colspan="2">0</td><td>1</td><td>PYPC</td></tr>
<tr><td>FAPC
S/F</td><td colspan="2">0</td><td>1</td><td></td><td></td><td>0</td><td>1</td><td></td><td>O/F
MP/F</td></tr>
<tr><td>PYPC
O/F
MP/F</td><td>1</td><td>0</td><td colspan="6"></td><td></td></tr>
<tr><td>T/F</td><td>0</td><td>1</td><td colspan="6"></td><td></td></tr>
</table>

Using the concept of location quotient, the locations of alternative clusters can be quantitatively described. In fact, the total standardized values of variables are equal to location quotients[2]. According to the order of location quotients for each variable, a weighting value is given to a cluster. Thus , the matrix of location quotients is obtained as shown in Table 6. Mean weighting value of a cluster can present its position in the system, this coinciding with the qualitative description.

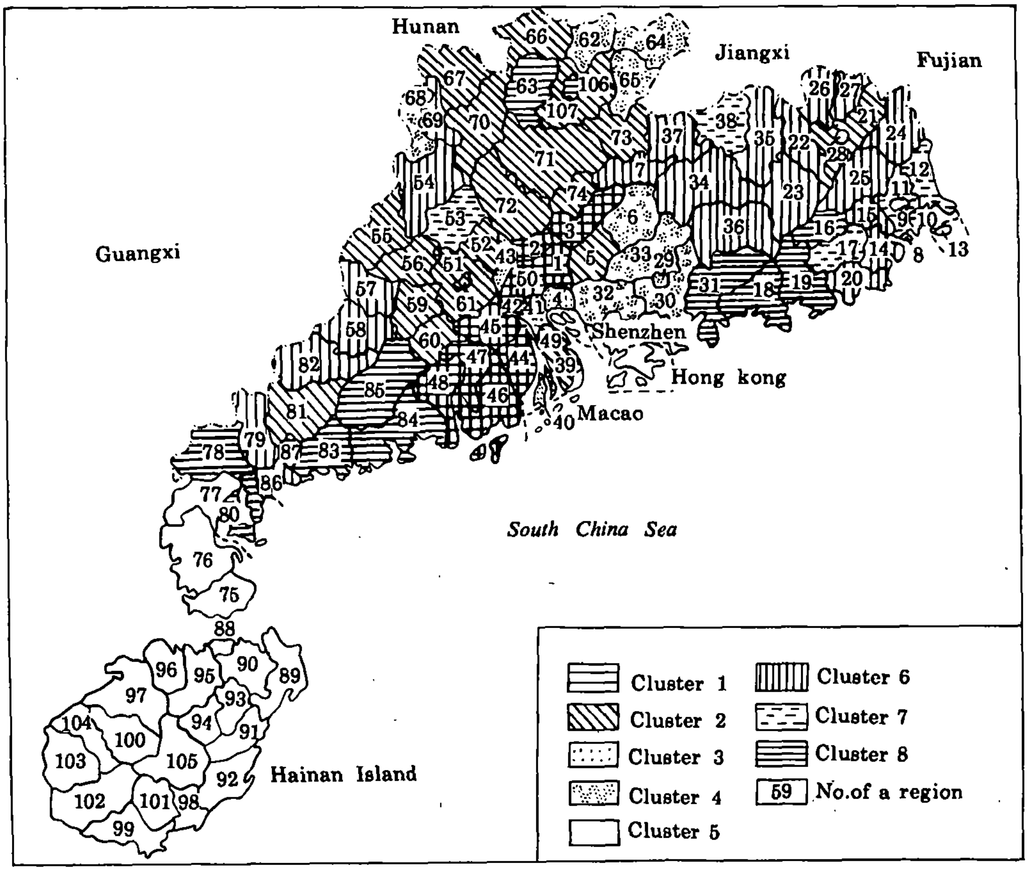

Fig. 4 The regional structure of agro—ecosystem in Guangdong

2. The Macroscopic Model of Agricultural Location

The von Thunen Theory, also called theory of agricultural location, was first suggested by von Thunen, a German economist (1826). The theory is an ideal location model of agricultural spatial distribution. Thunen used the distance to a market as a determinative factor forming complex agricultural landscape, and put other factors into constants according to the restrictive principle, so that he set up a famous distributive model of agricultural types that possesses a concentric structure centralized by a market[3].

This is an important concept—intensity in the von Thunen Theory, the optimal location of an agricultural production is determined by land income [4]. The shorter distance to

Table 6 Matrix of location quotients in 8 clusters

Variables	Cluster (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FAPC	5	4	1	3	2	7	8	6
GYPC	4	3	1	2	6	7	8	5
PYPC	3	6	2	1	8	7	5	4
JYPC	6	7	4	5	8	1	3	2
PHPC	1	4	3	7	3	8	6	5
R / G	1	3	1	2	5	4	6	5
S / F	3	6	1	5	2	8	7	4
T / F	4	3	7	6	8	2	1	5
O / F	2	5	7	3	8	4	1	6
MP / F	1	5	3	4	8	7	2	6
CF / F	2	3	2	7	8	4	5	6
mean weighting	2.91	4.45	2.73	4.09	6	5.36	4.72	4.91

a market leads to the higher efficiency of location, the more strength of land use, and the higher intensity. In the Thunen era , land and laborers were the most important input, and the intensity meant the value of laborer input on a unit of land. In modern agriculture, then, a large number of industrial auxiliary energy is input. Therefore, the input level of energy can present the intensity of an agricultural ecosystem.

Table 7 lists input levels of energy in eight clusters based on the sum of *MP / F* and *CF / F*. From Table 7, it can be seen that three high energy–input clusters (ie. cluster 1, 3 and 7) are located in two centeres of the system— the Northwest River Delta and the Chaoshan Plain. Four medium energy–input clusters (ie. cluster 2, 4, 6 and 8) form concentric circles around economic centeres. Even the small differences among these clusters coincide with their own locational characteristics, e.g. cluster 2 and 4. The lowest energy–input cluster 5 lies in the most distant location. This verifies macroscopically the von Thunen Theory, and does indeed give the spatial model of agricultural location in Guangdong.

Table 7 Input levels of energy in 8 clusters

Items	Cluster (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Average
MP / F+CF / F	0.39	0.24	0.35	0.20	0.09	0.22	0.31	0.19	0.25
input level	high	middle	high	middle	low	middle	high	middle	

REFERENCES

[1] 于崇文，数学地质的方法与应用，冶金工业出版社，1980.

[2] Gibson, A., Economic Geography, 1979.

[3] Singh,J., and Dhllon.S.S., Agricultural Geography, 1981.

[4] Grigg,D., An Introduction to Agricultural Geography, 1984.