第13卷第1期
1993年2月

水土保持通报
Bulletin of Soil and Water Conservation

Vol. 13 No. 1
Feb. ,1993

二维有序聚类及其在农业分区中的应用

——以陕西省延安地区为例

张 文 军

(西北农业大学·陕西杨陵·712100)

提 要

该文运用经作者修正的二维有序聚类方法,结合主成分分析和不同县市的农业资源指标分析,将延安地区划分为四个农业区域,针对各区的特点提出了治理措施,从而为该地区农业发展规划提供部分依据。同时,给出二维有序聚类方法的计算软件,为应用研究提供一类有效的工具。

关键词: 二维有序聚类 延安 农业分区

Two-dimensional Order Clustering and Its Application in the Regionalization of Agriculture

——Taking Yanan Prefecture in Shaanxi Province as an Example

Zhang Wenjun

(*Northwestern Agricultural University, Yangling, Shaanxi* 712100)

Abstract

Using the two-dimensional order clustering algorithm which had been revised by the author, and combining principal components analysis and resources indices analysis of agriculture in different counties and cities, Yanan prefecture is divided into four agricultural regions. The situation in northeastern region is the worst, and its developing strategies are to control soil erosion and to create satisfactory cycles for ecological system. The situation in northwestern region is also worse, and the developing strategies are to control soil erosion, to exploit underground water and to use the cultivars of crop and grass that perform well in cold climates. The situation in southwestern region is moderate, the potential for developing the southeastern region is the largest, its developing strategies are to increase stock raising, to exploit natural resources on the mountains, to enlarge the scale of crop production and to construct the base for agriculture, forestry and stock raising. These results provide a scientific basis for the developing planning of agriculture in this area. At the same time, the computer software of two—dimensional order clustering algorithm was given in order to provide a effect tool for the application and studies.

收稿日期:1992—06—22

Key words　two-dimensional order clustering　Yanan　regionalization of agriculture

由于气候条件差、水土流失严重、沟壑纵横等原因，延安地区的农业发展一直比较落后。随着开发热潮的兴起，迫切需要制定一套适应新形势的农业发展战略规划。无疑，水土保持农业资源的数值分区将有助于规划的科学性和合理性。

现有的聚类分析方法用于分区的最大缺点是，任何分法，均不能保证分区的毗连性，成为聚类分析应用上的障碍。而二维有序聚类既考虑分区的相似性，也兼顾分区的毗连性，尤其适用于地质学、地理学研究。本文运用经作者修正的二维有序聚类方法，用计算机软件进行了上述研究。

一、地区概况

延安地区位于陕西北部，面积 37 028.66km²，由 13 个县(市)组成。该地区属黄土高原丘陵沟壑区，平均沟壑密度 2.75km/km²，侵蚀模数 6 134t/(km²·a)，水土流失面积和耕地面积分别占总土地面积的 67.52%和 25.21%。农业人均耕地 10.4 亩，水浇地、川地等优质半优质耕地占耕地总面积的 13.27%，亩产粮 109.03kg。草场面积占总土地面积的 36.57%，亩产鲜草 275kg。森林覆盖率为 31.60%，人均活立木蓄积量 26m³。全区 80%保证率的降雨量为 392～513mm，农田蒸发量为 818.29mm。≥10℃积温为 2 724～3 863℃，最冷月平均气温－8.0～－4.7℃，负积温 520.75℃，无霜期 170.31 天。人均自产地表径流量 923.02m³，人均地下水可采量 45.70m³。除黄龙、甘泉两县外，其余 11 个县(市)均为黄河中游水土流失重点县(市)。

二、分区指标与二维有序聚类方法

(一)分区指标　农业资源分区既要考虑资源现状，又要考虑今后发展潜力，结合单 相关分析及地域特点，可筛选出 19 个分区指标：①沟壑密度(km/km²)，②水土流失面积占总土地面积比例(%)，③侵蚀模数(t/km²·a)，④耕地面积占总土地面积比例(%)，⑤农业人均耕地(亩)，⑥亩产粮(kg/亩)，⑦水浇地、川地等优质半优质耕地占耕地总面积比例(%)，⑧草场面积占总土地面积的比例(%)，⑨森林覆盖率(%)，⑩亩产鲜草量(kg/亩)，⑪人均活立木蓄积量(m³)，⑫80%保证率降水量(mm)，⑬农田蒸发量(mm)，⑭≥10℃积温(℃)，⑮负积温(℃)，⑯最冷月平均气温(℃)，⑰无霜期(d)，⑱人均自产地表径流量(m³)，以及⑲人均地下水可采量(m³)。13 个县(市)的 19 个指标数据整理自文献[1]～[4]。

(二)二维有序聚类　设平面上有 n 个点，每个点 m 个指标，记为

$$x_i = (x_{i1}, x_{i2})^T,$$

其中 $x_{i1} = (x_{i1}^1, x_{i1}^2)^T$，$x_{i2} = (x_{i2}^1, x_{i2}^2 \cdots\cdots x_{i2}^m)^T$，$i = 1,2,\cdots\cdots,n$，则聚类过程为：

1. 对点 $x_i, x_j, (j > i, i = 1,2,\cdots\cdots,n-1)$，构造聚类点集

$$G_1(i,j) = \{x_R, x_i \mid x_{R1}^2 > (x_{j1}^2 - x_{i1}^2)(x_{R1}^1 - x_{i1}^1)/(x_{j1}^1 - x_{i1}^1) + x_{i1}^2\},$$

$$G_2(i,j) = \{x_R, x_i \mid x_{R1}^2 < (x_{j1}^2 + x_{i1}^2)(x_{R1}^1 - x_{i1}^1)/(x_{j1}^1 - x_{i1}^1) + x_{i1}^2\};$$

和

$$\bar{G}_1(i,j) = \{x_R, x_i \mid x_{R1}^2 > (x_{R1}^2 - x_{i1}^2)/(x_{R1}^1 - x_{i1}^1)/(x_{j1}^1 - x_{i1}^1) + x_{i1}^2\},$$

$$\bar{G}_2(i,j) = \{x_R, x_i \mid x_{R1}^2 < (x_{j1}^2 - x_{i1}^2)(x_{R_1}^1 - x_{i1}^1)/(x_{j1}^1 - x_{i1}^1) + x_{i1}^2\}。$$

由此共得 $n(n-1)$ 种分法，对应的 $\{x_{i2}, 1 \leqslant i \leqslant n\}$ 也有 $n(n-1)$ 种分法。计算各自的误差函数

$$e[G(2)] = \sum_{L=1}^{2} D(i_L, i_{L+1} - 1),$$

其中，$D(i,j)=\sum_{L=i}^{2}(x_{L2}-\bar{x}_{ij})(x_{L2}-\bar{x}_{ij})$，

$$\bar{x}_{ij}=\frac{1}{j-i+1}\sum_{L=i}^{j}x_{L2},i<j,i=1,2,\cdots\cdots,n-1。$$

取 $min\ e[G(2)]$ 成立的一种分法，结果得两类。

表1 各原始指标的得分结果

原始指标	I	II	III	IV	V	VI	总得分
1	0.1497	0.2329	0.2122	−0.4654	0.1091	0.4104	0.9252
2	0.2891	0.1373	−0.1316	0.0443	0.2966	0.1972	0.9423
3	0.3185	0.0101	0.0631	−0.1391	0.2087	0.2137	0.9695
4	0.2772	0.2251	0.0078	0.0293	0.2676	0.0009	0.9508
5	0.2577	−0.2209	0.2254	0.1649	0.1737	−0.1199	0.9507
6	−0.2336	−0.0948	−0.3708	−0.3586	−0.0569	−0.0477	0.9369
7	−0.2515	0.0034	−0.2915	−0.2912	0.2904	0.1966	0.9196
8	0.0885	−0.2730	−0.1976	0.5158	0.0145	0.3423	0.9023
9	−0.3045	−0.1735	0.0219	−0.1247	−0.1141	0.0488	0.9579
10	−0.2860	0.0284	0.2927	−0.0387	−0.0721	−0.0876	0.8699
11	−0.2306	−0.1923	0.3248	−0.0303	−0.0510	0.4209	0.9273
12	−0.2775	−0.0518	0.0025	0.2544	0.4006	−0.0921	0.9217
13	0.0079	0.3873	0.2522	0.1328	−0.3663	0.0266	0.9408
14	−0.0136	0.4273	0.0598	0.1292	−0.2174	−0.0351	0.8858
15	0.2316	−0.2758	0.1449	−0.2727	−0.1250	−0.0523	0.9490
16	−0.2641	0.2275	−0.1834	0.2067	0.1356	0.2088	0.9880
17	−0.1929	0.3618	0.0155	0.0875	0.0621	0.2695	0.9554
18	−0.1588	−0.2463	0.4445	0.0912	0.1054	0.2812	0.9211
19	0.1813	−0.1457	−0.3317	0.0675	−0.5052	0.4229	0.9547
特征值	8.5029	4.41734	1.85243	1.35137	1.01054	0.63489	——

注：I——VI分别代表各综合指标

2. 设已分成 k 类，将它们分别用1.法试分为两类，选 $max\ min\ e[G(2)]$ 成立的一个进行分类，得到 $k+1$ 类。

3. 若 $k<n$，则继续分割。$k=n$ 时，分割结束，将全部过程画成聚类图。误差函数是分类个数的严格单调减函数，误差函数值为聚类距离。

三、农业资源区域划分

(一)主成分分析和综合 不同县(市)的19个分区指标是根据初选、单相关分析及反复比较取舍得到的。然而，这些指标间仍存在不同程度的相关性。无论定性分析还是定量分析，指标间的相关性会使数据信息得到不合理的重复作用，故可用主成分分析加以解决。作者将各县(市)19个指

标的数据在计算机上用主成分分析综合后得 6 个综合指标,从 19 个原始指标的得分结果(表 1)看,这 6 个指标负载的信息量占原 19 个指标所载信息量的 93.52%,是充分可信和可用的。也可用主成分分析结果在坐标图上划分区域。然而,该方法只相当于一般的聚类法,且信息损失通常较多(只用了前两个主成分)。例如,第 1、2 独立指标所载信息量只占 68.00%。因此,不宜选用此方法分区。

表 2　六个综合指标计算结果

地区	Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ	Ⅵ
延安市	0.92867	2.35874	0.22143	−0.01562	0.42569	−0.22268
延长县	−0.15556	5.47648	0.18435	2.16266	0.47055	−0.89055
延川县	6.31031	7.04965	0.26219	0.64136	−1.08925	0.88032
子长县	9.63058	4.25135	0.73635	−2.38793	0.40981	0.39053
安塞县	6.11546	0.99792	1.81792	−0.91846	1.11996	−0.47592
志丹县	8.20689	−7.49011	0.57055	0.21246	0.71353	−0.76279
吴旗县	13.50789	−6.30112	−1.58075	0.73158	−0.23981	0.77476
甘泉县	−3.23416	−3.41740	0.10131	−0.37422	−1.92325	−0.52279
富　县	−6.39920	−0.02615	−0.56167	−1.92790	−1.59661	−0.51707
洛川县	−2.71938	−0.11249	−4.21895	0.81091	0.36125	0.20179
宜川县	−5.63414	1.51025	0.96443	1.76248	−0.36698	−0.10769
黄龙县	−11.91129	−4.47454	3.41323	0.60939	0.26923	1.00119
黄陵县	−14.64608	0.17743	−1.91039	−1.30672	1.44587	0.25090

表 3　各县市的地理位置

地区	东经(度)	北纬(度)
延安市	110.0407	36.6053
延长县	110.0232	36.4913
延川县	110.0378	36.8603
子长县	109.6194	37.2458
安塞县	109.1503	36.9189
志丹县	108.6311	36.7781
吴旗县	108.0140	36.9833
甘泉县	109.1611	36.3254
富　县	109.1033	36.0624
洛川县	109.5335	35.7113
宜川县	110.1194	36.0519
黄龙县	109.9636	35.7181
黄陵县	108.9832	35.6747

从表 1 看,产生地域差异的主要因素是最冷月平均气温、侵蚀模数、森林覆盖率、无霜期、耕地占总土地面积比例及农业人均耕地等。这些因素也正是农业生产的基本资源因素或限制因素。

(二)二维有序聚类分区　作者编制了二维有序聚类通用软件(见附录)。为地理地质决策支持系统提供一种决策工具。应用该软件时,除读入不同县(市)6 个指标的数据外,首先需读入其经度、纬度(由该地中心的经纬度代表,结果整理于表 3)。

在计算机上得到二维有序聚类方法的聚类图如图 1 所示,根据实际需要加经验分析,确定将该地区分为四个区域(图 2)。子长县、安塞县、延川县、延安市、延长县和宜川县属第 I 区,为东北部农业区;吴旗县、志丹县、甘泉县、富县属第 II 区,为西北部农业区;黄陵县、洛川县属第 III 区,为西南部农业区;黄龙县属第 IV 区,为东南部农业区。

四、分区简述

根据分区结果计算的各区 19 个指标的平均值见表 3(指标含义见"分区指标"一节)。由该表可

综合分析四个区域的农业资源与发展情况。

(一)东北部农业区 本区农业资源状况很差,区内沟壑纵横,水土流失十分严重;广种薄收,林草稀少,降水量少,农田蒸发量大;冬季气候寒冷,人均自产地表水量少,资源开发后劲不足。今后要从治本入手,大规模种草种树,结合工程措施,大力减少水土流失,深度改良草场,选用抗旱高产粮草品种,启动生态良性循环。

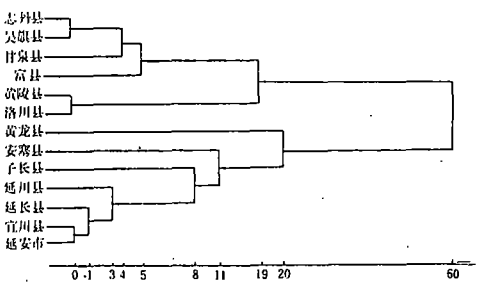

图1 二维有序聚类分区谱系图

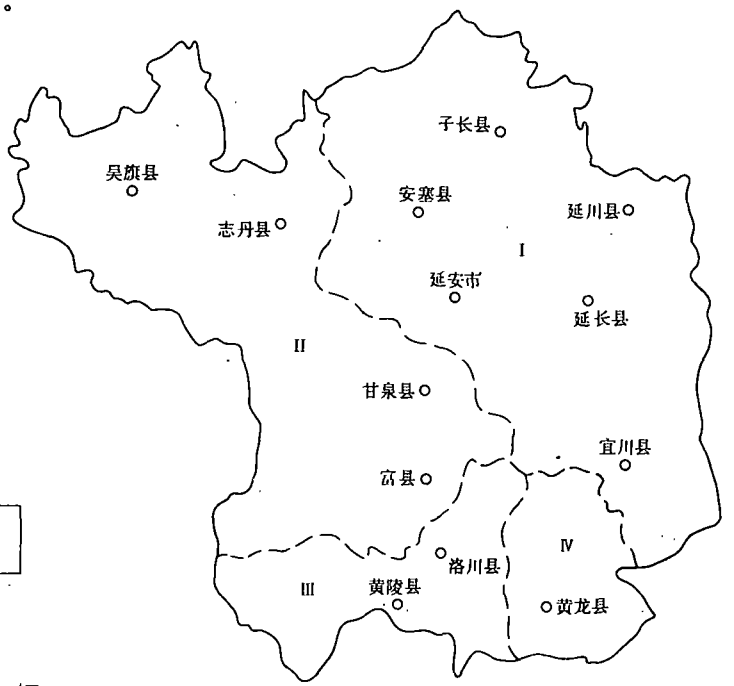

图2 延安地区农业资源分区图

表3 四个分区的指标值

区号	指标值			
	I	II	III	IV
1	3.7383	1.9550	1.7600	2.1600
2	79.7167	59.9500	62.4500	34.4000
3	7 857.1700	6 771.0000	1 045.0000	182.0000
4	34.4500	20.2500	18.8000	9.2000
5	10.6000	13.2000	4.9000	10.0000
6	90.6533	122.4675	175.4550	120.1900
7	12.4733	14.0525	40.5100	20.6900
8	29.3133	37.0925	34.1900	36.2600
9	18.0500	33.7750	49.2500	65.7000
10	231.9667	217.7000	252.5000	410.0000
11	7.6000	41.0000	44.5000	246.0000
12	441.8000	429.1500	483.2000	487.0000
13	843.5667	796.7425	789.1050	811.2300
14	3415.6833	3007.2750	3116.9500	2912.3000
15	429.1833	543.9000	339.9500	405.7000
16	−6.2667	−7.2250	−4.9500	−5.7000
17	177.6667	153.2500	179.5000	176.0000
18	854.5000	1105.5000	891.5000	2943.0000
19	42.6683	112.1275	57.5350	14.2800

(二)西北部农业区

本区农业资源状况较差,水土流失严重;草场面积大,产草量低,森林较为稀少;降水不足,气候相对偏冷,人均地下水可采量大。今后要采用生物和工程措施减少水土流失,改良草场,绿化荒山荒坡,在川道浅水区开采地下水源,培育生育期短和耐寒性好的粮食作物品种。

(三)西南部农业区

本区水土流失较轻,农牧作物产量高;降水量大,农田蒸发量小;林草面积大,气候相对温和,人均自产地表径流量小。今后除兼顾治理水土流失外,尚需进一步利用自然条件,培育水土保持涵养林,扩大优质农牧产品的生产和利用。

(四)东南部农业区

本区农业资源状况较好,水土流失很轻,粮草产量高;林草面积大,人均活立木量大;降雨多,人均自产地表径流量大,气候相对温和。今后应在山区发展林牧业,以此推动农业发展;开发山区资源;充分利用地表径流,在川道发展粮食作物;建立牧林草产品创收基地。

五、讨论

1. 区位指标的选择方法灵活多样，就经、纬度指标而言，可用的方法是选择地域中心经纬度、地域代表性经纬度以及增大取样点个数等等。所研究的地域愈大，区位值对结果的影响愈小。但无论何种方法，均须由聚类结果的正确性来验证。本研究的结果是符合延安地区实际的。

2. 经典的聚类方法多种多样，但其共同特征是没有考虑地域分类中的毗连性问题。最优分割法是一维毗连聚类法，其有效性已得到广泛承认。二维有序聚类就是在二维方向上综合最优分割过程的产物，从而克服了传统聚类方法的缺点。所以，对于以地域性研究为特点的地理、地质学有独到用途，可进一步探索其应用途径。

附录　二维有序聚类应用软件

```
100  READ N,M
120  DIM X(M+2,N),E(N*(N-1)/2,2),Q(N,N*
     (N-1)/2),P(M+2),U(M+2)
125  DIM Y(N*(N-1)/2,N),Z(N*(N-1)/2),C(N
     *(N-1)/2)
130  FOR I=1 TO M+2
140  FOR J=1 TO N
150  READ X(I,J)
160  NEXT J
170  NEXT I
180  FOR I=3 TO M+2
190  U(I)=0
200  P(I)=0
210  FOR J=1 TO N
220  P(I)=P(I)+X(I,J)
230  NEXT J
240  P(I)=P(I)/N
250  FOR K=1 TO N
260  U(I)=U(I)+(X(I,K)-P(I)) ∧2
270  NEXT K
280  U(I)=SQR(U(I)/(N-1))
290  FOR J=1 TO N
300  X(I,J)=(X(I,J)-P(I)/U(I)
310  NEXT J
320  NEXT I
340  FOR K=1 TO N
350  Y(O,K)=K
360  NEXT K
362  FOR A=1 TO N*(N-1)/2
364  C(A)=0:Z(A)=0
366  NEXT A
370  A=0:G=0
380  GOSUB 700
385  SD=A
390  G=G+2
400  A=G
410  GOSUB 1300
416  FOR A=1 TO G
417  IF C(A)>1 THEN GOTO 420
418  NEXT A
419  END
420  A=1
430  IF C(A)=0  OR  C(A)=1 THEN  GOTO 470
440  N=C(A)
450  GOSUB 700
460  Z(A)=Z
470  A=A+1
472  IF A=G+1 THEN GOTO 480
475  GOTO 430
480  HF=-1E+08
490  FOR A=1 TO G
495  IF C(A)=0 THEN GOTO 510
500  IF Z(A)>HF THEN LET HF=Z(A)
510  NEXT A
520  FOR A=1 TO G
525  IF C(A)=0 OR C(A)=1 THEN GOTO 570
530  IF Z(A)<>HF THEN GOTO 570
540  N=C(A)
550  C(A)=0
560  GOTO 380
```

```
570 NEXT A
580 END
700 B=0
780 FOR I=1 TO N-1
800 FOR J=I+1 TO N
820 B=B+1
830 R=Y(A,I) : S=Y(A,J)
832 Q(Y(A,K),B)=0
833 NEXT K
835 V=1
837 E(B,V)=0
838 D=0 : F=0
840 FOR L=3 TO M+2
842 P(L)=0 : U(L)=0
844 NEXT L
846 FOR K=1 TO N
890 IF Y(A,K)=R THEN GOTO 920
900 IF Y(A,K)=S THEN GOTO 970
902 IF X(1,Y(A,J))=X(1,Y(A,I)) THEN GOTO
    914
908 W=(X(2,Y(A,J))-X(2,Y(A,I)))*(X(1,Y
    (A,K))-X(1,Y(A,I)))/(X(1,Y(A,J))-X(1,
    Y(A,I)))+X(2,Y(A,I))
910 IF X(2,Y(A,K))>W THEN GOTO 970
912 GOTO 920
914 IF X(1,Y(A,K))<X(1,Y(A,I)) THEN GOTO
    970
930 FOR L=3 TO M+2
940 P(L)=P(L)+X(L,Y(A,K))
950 NEXT L
960 GOTO 1020
970 F=F+1
980 Q(Y(A,K),B)=V
990 FOR L=3 TO M+2
1000 U(L)=U(L)+X(L,Y(A,K))
1010 NEXT L
1020 NEXT K
1030 FOR L=3 TO M+2
1040 P(L)=P(L)/D
1050 U(L)=U(L)/F
1060 NEXT L
1070 FOR K=1 TO N
1080 FOR L=3 TO M+2
1090 IF Q(Y(A,K),B)=V THEN GOTO 1120
1100 E(B,V)=E(B,V)+(X(L,Y(A,K))-P(L))^
     2
1110 GOTO 1130
1120 E(B,V)=E(B,V)+(X(L,Y(A,K))-U(L))^2
1130 NEXT L
1140 NEXT K
1150 R=Y(A,J) : S=Y(A,I)
1160 IF V=2 THEN GOTO 1190
1170 V=V+1
1180 GOTO 837
1190 NEXT J
1200 NEXT I
1210 Z=1E+09
1220 FOR B=1 TO N*(N-1)/2
1230 FOR V=1 TO 2
1240 IF E(B,V)<Z THEN LET Z=E(B,V)
1250 NEXT V
1260 NEXT B
1280 RETURN
1300 PRINT"e(G(2))=";Z;" ";"(";
1310 C=0 : H=0
1320 FOR B=1 TO N*(N-1)/2
1330 FOR V=1 TO 2
1340 IF E(B,V)<>Z THEN GOTO 1490
1350 FOR K=1 TO N
1360 IF Q(Y(SD,K),B)()V THEN GOTO 1375
1370 GOTO 1390
1375 C=C+1
1376 Y(A-1,C)=Y(SD,K) : PRINT Y(SD,K);"";
1390 NEXT K
1400 PRINT ")";" ";"""("
1410 FOR K=1 TO N
1420 IF Q(Y(SD,K),B)=V THEN GOTO 1440
1430 GOTO 1460
1440 H=H+1
1445 Y(A,H)=Y(SD,K):PRINT Y(SD,K);"";
1460 NEXT K
1470 PRINT ")"
1480 GOTO 1510
1490 NEXT V
1500 NEXT B
1510 C(A-1)=C:C(A)=H
1520 RETURN
```

软件说明：

M 和 N 分别为指标数和样品数。130—170 句输入原始数据，先输经度和纬度，后输指标数据。1300 句打印误差函数值，即聚类距离。1376 句和 1445 句打印分类结果。

参 考 文 献

〔1〕陕西省农业区划委员会办公室等. 陕西农业地图册. 西安. 地图出版社，1988 年
〔2〕陕西省农牧厅等. 陕西省种植业资源与区划. 西安：陕西科学技术出版社，1987 年
〔3〕陕西师范大学地理系《延安地区地理志》编写组. 延安地区地理志. 西安：陕西人民出版社，1983 年
〔4〕西北大学地理系《陕西农业地理》编写组. 陕西农业地理. 西安：陕西人民出版社，1979 年

（上接第 33 页）

本文在野外资料收集和成文过程中，曾得到林文棟教授大力协助和精心指导。对此，笔者表示衷心感谢！

参 考 文 献

〔1〕陈邦本等著.《江苏海岸带土壤》. 河海大学出版社，1988 年
〔2〕北京林学院主编.《数理统计》. 北京：中国林业出版社，1985 年