Korean J. Ecol. 10(2) : 91~107, 1987

植被 Data에 對한 Information과 Entropy 理論의 實用硏究

朴 勝 太
(全北大 師大 生物敎育科)

A Study on the Presentation of Idea in Information and Entropy Theory in Vegetation Data

Park, Seung Tai
(*Dept. of Biology Educ., Jeonbug Nat'l. Univ.*)

ABSTRACT

This study is concerned with some methods and applications, used as a basis on information and entropy analysis of vegetation data.

These methods are adopted for the evaluating the effect of sampling intensity on information, which represents the departure of observed variable from standard component. Classes on the data matrix are calculated by using marginal dispersion array for rank and weighting information program. Finally the information and entropy are computed by applying seven options.

On the application of vegetation studies, two models for cluster analysis and analysis of concentration are explained in detail. Cluster analysis is based on use of equivocation information and Rajski's metrics. The analysis of concentration utilizes coherence coefficience being transformed values, which has been adjusted from blocks and entropy values.

The relationship between three vegetation clusters and four stands of Naejangsan data is highly significant in 79% of total variance. Cluster A relatively tends to prefer north side, and cluster C south side.

結 論

生態學의 複雜한 資料(multidimensional data)에 대하여 information理論을 適用하는 것은 modeling을 하거나 統計的인 資料의 單純化(simplification) 및 資料 變數에 의한 傾向性(trend)을 찾는 것과는 상당한 차이가 있다(Legendre and Legendre, 1983; Pielou, 1977; Gauch, 1982; Orloci, 1978).

生態學的 硏究에서 種의 複雜한 關係, 層化(stratification) 및 季節的인 變化등에 대한 時間的 또는 空間的인 變化를 定量的으로 敍述할 때 硏究條件에 따라 個體數 또는 生體量을 나타내는 evenness와 richness에 의해 類型分析(pattern analysis), 區分分析(discrimination

* 이 論文은 85年 尖端科學 技術分野 硏究費 支援에 의한 것임.

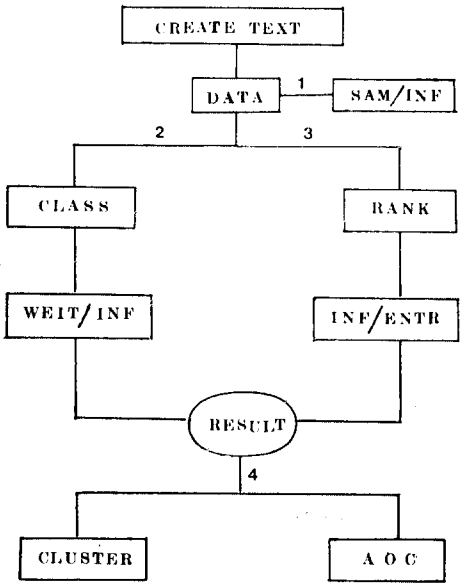

Fig. 1. Flowchart of information/entropy analysis. SAM/INF=sampling information; CLASS=dispersion array program; RANK=rank program; WEIT/INF=weighting information program; INF/ENTR=information and entropy program; AOC=analysis of concentration.

analysis) 및 豫見分析(predictive analysis)을 하게 된다(Greig-Smith, 1957; Orloci, 1972; MacArthur, 1965; Feoli *et al.*, 1984; Shannon, 1948).

Information理論은 意思傳達理論(communication theory)을 基礎로 하여 發達된 것으로 entropy 分析과 divergence 分析으로 區分되며, Renyi(1961)는 entropy에 대하여 disorder量으로 H^{α}를 써서 α의 次數에 따라 Shannon의 entropy函數와 Simpson指數를 算出할 수 있게 했으며 Kullback(1959)은 divergence에 대하여 2(Information)=χ^2로 나타내어 最小識別統計(minimum discrimination information statistics)의 基礎를 마련하므로써 資料에서 期待値와 觀側値와의 關係에 대한 divergence를 산출하여 分散分析(dispersion analysis), 多樣性分析(diversity analysis) 및 豫見分析(predictive analysis) 등을 가능하게 했다(Fisher, 1948; Margalef, 1958; McIntosh, 1967; Orloci, 1972).

본 硏究에서는 Fig. 1에서와 같이 information/entropy理論을 基礎로 하여 ① Data sampling의 效率의 計算을 위하여 最適標集의 intensity(n)를 정하고 이에 따른 quadrat數와 크기를 결정할 수 있게하고 ② weighting information을 할때 資料의 階級(class)을 임의로 조절하여 값을 算出하고 ③ 방대한 資料에 대하여 RANK program으로 rank를 정하여 줄일 수 있게 하고 ④ 計算된 information 값을 이용하여 結果를 解析할 수 있도록 clustering 方法과 集中度分析(AOC)으로 information/entropy값을 이용하므로서 生態學 資料分析 및 解析을 용이하게 할 수 있게 한 것이다.

方 法

Information理論 Information은 意思傳達理論을 기초로 하여 발달되었으며 Fisher(1948)는 logarithm 函數를 이용하여 確率을 information (I)으로 나타 냈다.

$$I(E)=-\ln P(E):\ E=1, 2, \cdots, s \quad \cdots\cdots(1)$$

이때 事象(E)은 s번 試行할 수 있으며 information 값은 항상 +값(positive)이며 다음과 같이 變形할 수 있다.

$$I(E)=-2\sum_{E=1}^{S}\ln \Pi(E) \quad \cdots\cdots(2)$$

또한 Kullback(1959)은 카이자승(χ^2)變量으로 이를 변형시켜 information과 χ^2 關係를 究明하였다.

$$\chi^2=2I \quad \cdots\cdots(3)$$

Information은 entropy와 divergence로 區分할 수 있는데(Orloci, 1978), entropy는 어떤 分

布에 대한 disorder를 logarithm 함수로 표시했고, Renyi(1961)는 disorder (H^α)를 다음과 같이 나타냈다.

$$H^\alpha=\frac{\ln\sum_{E=1}^{S}P(E)^\alpha}{1-\alpha} \quad \cdots\cdots(4)$$

위의 식 (4)에서 $\alpha=0$일때 $H^0=\ln S$로 Hmax가 되며, $\alpha\rightarrow 1$ 일때 L'Hopital 法則을 이용하여 풀면 $H^1=-\sum_{E=1}^{S}P(E)\ln P(E)$가 되는데 이는 Shannon index이며, $\alpha=2$일때 $H^2=-\ln\sum_{E=1}^{S}P(E)^2$로 Simpson index이다.

또한 entropy는 joint entropy ($I(\boldsymbol{a},\boldsymbol{b})$), mutual entropy ($I(\boldsymbol{a};\boldsymbol{b})$) 및 equivocation entropy ($E(\boldsymbol{a};\boldsymbol{b})$)등으로 區分되는데 이들의 計算은 다음 program에서 자세히 설명된다.

Divergence는 觀側値와 期待値에 대한 確率이나 頻度를 이용하여 Kullback(1959)는 다음과 같이 나타냈다.

$$\chi^2=2I(\boldsymbol{P};\boldsymbol{P}^0) \quad \cdots\cdots(5)$$

위의 식 (5)에서 觀側値의 確率은 $P=\{P(1)P(2)\cdots\cdots P(s)\}$이고 期待値의 確率은 $P^0=\{P^0(1)\ P^0(2)\cdots\cdots P^0(s)\}$이며 divergence를 disorder (H)로 나타낼 때 다음과 같다.

$$H(\boldsymbol{P};\ \boldsymbol{P}^0)=H(\boldsymbol{P}^0)-H(\boldsymbol{P})$$

$$=\sum_{E=1}^{S}P(E)\ln P(E)/P^0(E) \quad \cdots\cdots(6)$$

또한 頻度(F)를 이용할 때에도 觀側頻度($\boldsymbol{F}$)와 期待頻度로($\boldsymbol{F}^0$) 計算할 수 있다.

$$\chi^2=2I(\boldsymbol{F};\ \boldsymbol{F}^0) \quad \cdots\cdots(7)$$

이때 $\boldsymbol{F}=n\boldsymbol{P}$이며 $\boldsymbol{F}^0=n\boldsymbol{P}^0$로서 n은 標集크기이며 期待頻度($\boldsymbol{F}^0$)의 計算은 對應되는 두 頻度의 平均을 이용하여 算出한다.

그러나 觀側頻度(f_{ij} 또는 f_{hj})만을 이용하여 對應되는 두 quadrat間의 divergence를 計算할 경우는 다음과 같다.

$$I(\boldsymbol{A};\ \boldsymbol{B})=I\ (\text{quadrat1};\ \text{quadrat2})$$

$$=\sum_h\sum_j f_{hj}\ln 2f_{ij}/(f_{h1}+f_{h2})$$

$$h=1,\cdots\cdots,s;\ j=1,2 \quad \cdots\cdots(8)$$

이때 $\boldsymbol{A}$와 $\boldsymbol{B}$는 quadrat이며 group으로 區分된 quadrat에서도 計算할 수 있다.

또한 種과 quadrat의 information 計算은 다음과 같이 할 수 있다.

$$I\ (\text{species};\ \text{quadrat})=\sum_h\sum_j f_{hj}\ln f_{hj}f../(f_{h\cdot}f_{\cdot j})$$

$$h=1,\cdots,s;\ j=1,2 \quad \cdots\cdots(9)$$

위의 식(9)에서 $f..=\sum_h\sum_j f_{hj}$이고 $f_{h\cdot}=\sum_j f_{hj}$이며 $f_{\cdot j}=\sum_h f_{hj}$이다.

이때 식(8)과 식(9)의 차이 값을 divergence (Δ)로 이용한다.

$$\Delta(1,2)=I(\boldsymbol{A};\ \boldsymbol{B})-I(\text{species};\ \text{quadrat})$$

$$=\sum f_{\cdot j}\ln 2f_{\cdot j}/f..$$

$$j=1,2 \quad \cdots\cdots(10)$$

본 硏究에서는 information/entropy 理論만을 이용하여 植被(vegetation) 資料에 적용할 수 있도록 program 했으며 entropy 값으로 clustering 하거나 集中度(concentration)를 算出하여 解析할 수 있게 했다.

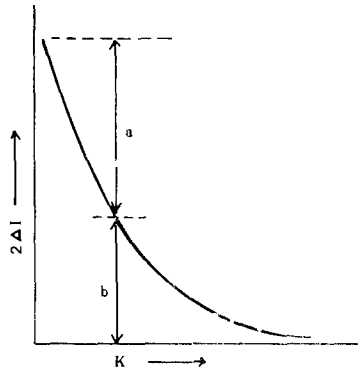

Fig. 2. Relationship of sampling effort and information. The vertical scale indicates information loss due to not counting entries beyond a maximum K. The symbol a indicates the gain of information and b information loss.

Program ① Sampling information program(SAM/INF); 標集方法은 硏究戰略과 硏究條件에 따라 quadrat 크기와 수를 한정해야하며 특히 quadrat 수는 結果에 상당한 영향을 하게 된다(Greig-Smith, 1957). 본 硏究에서는 Fig. 2와 같이 sampling effort에 따라 information 得失에 차이가 있으므로 quadrat의 적합한 數와 quadrat 크기에 한정된 intensity(n)의 결정은 information loss ($2\Delta I$)가 最小일 때를 다음과 같은 식으로 정했다.

$$2\Delta I=2\sum_{i=1}^{r}\sum_{j=1}^{c} f_{ij}\ln f_{ij}/k+(n-f_{ij})\ln(n-f_{ij})/(n-k)$$
$$i=1,2,\cdots,r;\ j=1,2,\cdots,c \quad\cdots(11)$$

위의 식(11)에서 species는 r, quadrat는 c인 資料에서 觀側된 要素인 f_{ij} 및 $(n-f_{ij})$와 期待된 要素 k 및 $(n-k)$의 거리 계산은 $f_{ij}\leqq k$ 일때만 가능하다. n은 한개 quadrat내의 最小數(intensity)이며 k는 quadrat 最適數로 intensity에 따라 quadrat 크기를 결정하게 된다(Orloci, 1970).

② Class program(CLASS); 植被資料를 分析할 때 array하는 方法이 4가지가 있다. R-array는 $r\times c$ matrix 資料에서 r變數(predictor)의 分散配列(dispersion array)을 말하며, Q-array는 c變數(object)를 分散配列하는 것이다. Predictive array는 $r\times c$ matrix에서 resemblance를 산출하여 $r\times r$ 또는 $c\times c$ array하는 것이고 diversity array는 R 또는 Q-array하는 방법을 이용하여 多樣度를 산출하는 array法이다(Feoli *et al.*, 1984).

array 할때 class의 限界를 임의로 정하고(본 연구에서는 6개 class; Table 4와5 참고) marginal dispersion을 산출하여 weighting information을 구하거나 RANK program 을 운영할 때 이용했다.

③ Rank program (RANK); information에 의한 變數의 rank를 정하는 理論은 다양하나 (Orloci, 1976) 본 硏究에서는 equivocation information ($E(\boldsymbol{a};\boldsymbol{b})$)을 이용하여 rank를 결정하고 資料變數를 줄였다.

④ Weighting information program(WEIT/INF); Orloci(1978)의 program을 이용하였다.

⑤ Information entropy program(INF/ENTROPY); 生態學 資料分析에 대하여 Fig. 3과 같이 venn diagram에서 7가지 information을 계산할 수 있도록 program하고 각각을 선택할 수 있게했다.

① information a; $I(\boldsymbol{a})$

$$I(\boldsymbol{a})=-\sum_j f_{hj}\ln f_{hj}/f_{h\cdot}$$
$$=\sum_j f_{hj}\ln f_{h\cdot}-\sum_j f_{hj}\ln f_{hj}$$
$$j=1,2,\cdots,S_h \quad\cdots\cdots\cdots\cdots(12)$$

② information b; $I(\boldsymbol{b})$

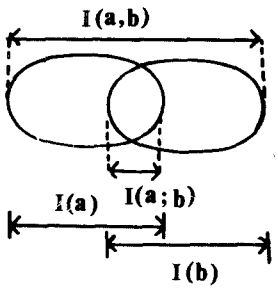

Fig. 3. Venn diagram.

$$I(\boldsymbol{b})=-\sum_j f_{ij}\ln f_{ij}/f_{i\cdot}$$

$$=\sum_j f_{ij}\ln f_{i\cdot}-\sum_j f_{ij}\ln f_{ij}$$

$$j=1,2,\cdots,S_i \quad \cdots\cdots(13)$$

③ joint information; $I(\boldsymbol{a},\boldsymbol{b})$

$$I(\boldsymbol{a},\boldsymbol{b})=-\sum_j\sum_k f_{hj,ik}\ln f_{hj,ik}/f_{h\cdot,i\cdot}$$

$$=\sum_j\sum_k f_{hj,ik}\ln f_{h\cdot,i\cdot}-\sum_j\sum_k f_{hj,ik}\ln f_{hj,ik}$$

$$j=1,\cdots,S_h;\ k=1,\cdots,S_i \quad \cdots\cdots(14)$$

④ mutual information (interaction information); $I(\boldsymbol{a};\boldsymbol{b})$

$$I(\boldsymbol{a};\boldsymbol{b})=I(\boldsymbol{a})+I(\boldsymbol{b})-I(\boldsymbol{a},\boldsymbol{b})$$

$$=\sum_j\sum_k f_{hj,ik}\ln f_{hj,ik}f_{h\cdot,i\cdot}/f_{hj,i\cdot}f_{h\cdot ik}$$

$$j=1,\cdots,S_h;\ k=1,\cdots,S_i \quad \cdots\cdots(15)$$

⑤ equivocation information; $E(\boldsymbol{a};\boldsymbol{b})$

$$E(\boldsymbol{a};\boldsymbol{b})=I(\boldsymbol{a},\boldsymbol{b})-I(\boldsymbol{a};\boldsymbol{b}) \quad \cdots\cdots(16)$$

⑥ Rajski metrics; $d(\boldsymbol{a};\boldsymbol{b})$

$$d(\boldsymbol{a};\boldsymbol{b})=\frac{E(\boldsymbol{a};\boldsymbol{b})}{I(\boldsymbol{a},\boldsymbol{b})} \quad \cdots\cdots(17)$$

⑦ coherence coefficient; $r(\boldsymbol{a};\boldsymbol{b})$

$$r(\boldsymbol{a};\boldsymbol{b})=|1-d^2(\boldsymbol{a};\boldsymbol{b})|^{1/2} \quad \cdots\cdots(18)$$

모든 program은 APPLE BASIC, DOS 3.3을 이용할 수 있게 제작 되었다.

이때의 $\boldsymbol{a},\boldsymbol{b}$는 대립되는 變數로서 빈도나 밀도를 나타낸다.

Clustering(CLUSTER) Information에 의한 Clustering은 equivocation information과 Rajski metrics를 이용하여 sum of square(SS) algorithm으로 fusion하였고 cluster의 결정은 information SS값을 임으로 선정하여 구분 하였다. Clustering 效率은 區分된 cluster內의 SS(W) information을 計算하고 全體 cluster SS(T) information에서 빼면 cluster間의 SS(B) information이 된다. 이를 이용하여 $\frac{B}{T}\times 100(\%)$로 效率을 計算 했다(Sneath and Sokal, 1973).

集中度分析(AOC) 植被內에 分布된 種과 環境과의 關係를 正準分析法(canonical analysis)을 이용하여 傾向을 추출하는 方法인데 본 硏究에서는 coherence coefficient($r(\boldsymbol{a};\boldsymbol{b})$) 값($\boldsymbol{I}_{pp}$)을 이용하여 資料를 다음과 같이 adjust ($\boldsymbol{A}_{pq}$) 했고 N은 식(20)과 같이 補正값이다.

$$\boldsymbol{A}_{pq}=\boldsymbol{I}_{pp}R_{pq}/N \quad \cdots\cdots(19)$$

$$N=pq/|p+q|^{1/2} \quad \cdots\cdots(20)$$

위의 식(19)에서 $\boldsymbol{R}_{pq}$는 raw data이며 $p=1,\cdots,r$로서 data matrix에서 種을 나타내며 $q=1,\cdots,c$는 quadrat를 나타낸다. Clustering 結果에 따른 種의 cluster와 調査地所를 정리하여 分割表를 Block으로 만들어서 이에 대한 集中度分析은 AOC program으로 내장산 資料(Table 2) 만을 分析 하였다(Park, 1984).

資料標集 카나다 Sifton Bog 中心地域에 인접된 곳에 300 區域을 정하고 random으로 (0.5×0.5)m 크기의 quadrat 30個를 設置하고 出現한 植物 頻度를 조사하여 37種×30 quadrat의 data를 얻었다(Orloci, 1970) 이를 RANK program으로 20種×30 quadrat로 줄여서 資料로 이용 했다(Table 1).

Table 1. Sifton Bog data (Orloci, 1970)

Species	Quadrat number																														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1. *Sphagnum fuscum*	12	13	10	23	14	0	13	0	13	24	21	17	9	18	19	19	16	23	22	20	8	9	5	17	0	12	0	0	13	21
2. *Sphagnum palustre*	1	18	16	4	5	0	8	0	5	5	5	10	14	3	14	4	4	4	9	12	18	17	21	5	0	18	0	0	11	2
3. *Vaccinium oxycoccus*	2	4	4	7	15	0	8	0	10	17	12	13	5	10	1	9	0	9	4	1	1	7	13	15	0	3	0	0	0	10
4. *Rhamnus frangula*	1	0	8	1	2	11	0	14	1	3	3	0	2	0	16	0	3	0	0	6	6	3	0	1	12	9	21	1	0	0
5. *Rhynchospora alba*	11	0	0	11	11	0	0	0	8	4	14	8	2	2	0	15	0	20	3	0	0	0	0	19	0	13	0	0	0	19
6. *Chamaedaphne calyculata*	25	21	22	25	19	0	23	0	20	22	18	25	21	25	1	20	16	20	25	16	18	19	23	22	0	5	0	0	25	25
7. *Andromeda glaucophylla*	0	0	0	3	8	0	7	0	6	12	6	0	0	0	0	4	0	11	1	2	0	0	0	7	0	0	0	0	0	0
8. *Drosera rotundifolia*	5	1	2	2	9	0	3	0	5	12	6	1	1	3	0	4	1	6	5	0	1	2	2	1	0	0	0	0	1	2
9. *Sphagnum recurvum*	6	0	0	0	0	0	2	0	5	0	5	0	2	4	0	2	0	0	0	2	0	0	0	2	0	6	0	0	5	0
10. *Hypericum verginianum*	0	0	0	3	0	0	0	0	0	5	9	0	3	0	0	3	0	0	0	0	7	12	0	0	0	0	0	0	1	0
11. *Thelypteris palustris*	0	0	0	0	0	3	0	5	0	0	0	1	0	0	0	0	0	0	0	0	8	16	1	0	0	0	1	0	2	0
12. *Sphagnum riparium*	10	2	1	0	0	0	0	0	4	1	0	1	6	1	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0
13. *Picea mariana*	0	5	1	1	12	1	4	0	0	5	0	2	1	1	3	3	3	1	2	4	0	0	1	3	0	4	0	0	0	2
14. *Acer rubrum*	4	7	1	3	1	5	2	5	0	3	3	4	5	3	9	3	4	2	1	1	1	0	1	10	4	7	3	5	0	0
15. *Carex sp*	1	0	0	0	3	0	0	0	16	3	0	2	5	0	0	5	0	2	1	0	0	0	0	0	0	0	0	0	0	3
16. *Pogonia ophioglssoides*	0	0	0	0	0	6	0	0	0	0	12	0	3	0	0	0	0	0	0	0	1	1	2	2	0	0	0	0	0	0
17. *Impatiens biflora*	0	0	0	0	0	13	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	5	22	0	0
18. *Solanum dulcamara*	0	0	0	0	0	4	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0	3	6	0	0
19. *Sphagnum capillaceum*	0	0	0	1	9	0	5	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	3	5	0	0	0	0	0	0
20. *Sarracenia purpurea*	0	0	0	0	4	0	1	0	1	2	0	0	0	0	0	0	0	1	3	1	0	0	3	0	0	0	0	0	0	0

Table 2. Naejangsan data

Species	Quadrat number																															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1. *Lespedeza bicolor*	0	0	2	0	10	10	0	0	1	0	2	0	0	10	10	15	15	15	20	20	16	16	10	4	4	4	10	20	20	20	16	16
2. *Lespedeza maximowicz*	0	0	0	8	10	14	4	4	4	8	8	8	6	1	0	0	1	0	8	8	14	8	8	0	1	1	1	14	14	20	20	14
3. *Sasa purpurascens*	24	14	10	10	14	20	8	10	8	14	14	14	8	18	18	4	6	4	8	8	16	16	22	25	20	20	24	20	18	20	20	20
4. *Fraxinus sieboldiana*	0	0	0	8	8	8	0	0	0	0	0	0	4	8	8	6	6	6	14	14	6	0	4	8	4	0	0	14	14	18	18	20
5. *Pueraria thunbergiana*	18	4	8	6	0	0	8	8	8	14	14	14	8	8	6	1	8	0	0	0	1	14	14	8	8	0	0	14	14	14	0	0
6. *Cornus controversa*	18	18	18	18	18	18	18	14	18	14	18	8	8	8	8	14	16	14	4	4	14	14	20	14	20	18	18	8	8	4	4	0
7. *Daphniphyllum macropodum*	0	0	0	0	25	14	14	14	20	8	14	4	8	10	10	0	0	0	0	0	1	4	20	0	0	0	0	0	0	0	0	0
8. *Acer japonica*	24	25	20	11	22	22	24	20	14	14	14	12	14	14	14	14	14	14	14	14	22	22	22	22	20	20	20	18	18	8	8	8
9. *Carpinus laxiflora*	0	0	0	5	8	8	0	0	0	0	0	20	10	0	0	5	8	0	0	8	4	2	2	0	0	1	0	0	6	2	1	1
10. *Akebia quinata*	4	1	1	0	8	0	1	1	1	0	2	2	1	0	0	0	0	0	0	0	4	8	8	6	0	4	0	18	12	4	0	0
11. *Pinus densiflora*	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	16	20	24	22	24	0	0	0	20	20	18	18	16	22	18	14	14
12. *Rhododendron schlippenbachii*	0	0	0	0	0	0	0	0	0	0	0	0	0	6	16	18	14	14	8	8	8	8	4	0	0	0	0	0	2	0	0	4
13. *Actinidia arguta*	10	8	8	4	0	0	8	4	8	8	8	8	8	8	6	0	2	0	0	1	0	0	0	0	4	4	2	8	6	6	4	0
14. *Staphylea bumalda*	10	0	0	0	0	0	4	4	0	8	8	4	0	4	2	0	6	0	0	1	0	0	1	0	1	0	1	0	0	0	0	0
15. *Carpinus tschonoskii*	12	8	10	22	12	14	0	0	0	24	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16. *Celastrus orbiculalus*	8	0	1	0	0	0	0	1	0	0	1	0	0	8	8	0	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0
17. *Hydrangea serrata*	4	2	6	2	0	0	0	0	0	4	2	2	1	0	0	0	0	0	1	0	0	0	0	0	0	2	1	4	0	1	0	0
18. *Zelkova serrata*	4	1	0	0	0	0	14	0	0	0	14	8	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
19. *Styrax japonica*	8	0	8	0	0	0	1	0	0	0	0	0	0	2	4	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0
20. *Smilax china*	0	1	4	4	1	0	0	1	2	4	4	4	4	0	0	0	0	2	1	0	0	1	0	0	0	2	0	0	1	4	0	0
21. *Acer mono*	2	0	0	2	1	1	0	1	0	2	2	0	2	2	0	0	2	2	0	2	0	0	0	2	4	0	1	0	2	1	0	0
22. *Rhus javanica*	0	1	0	4	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0
23. *Benzoin obtusilobum*	0	4	1	1	0	0	4	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24. *Meliosma myriantha*	0	0	0	0	0	0	5	25	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0

또한 우리나라에서 內藏山을 國立公園으로 지정할 때 조사된 자료를 토대로(Park, 1974) 1985년 4월부터 10월까지 내장산 일대에서 4個地域으로 區分하여 전망대 부근의 6個地所, 선인봉 부근의 9個地所, 까치봉 부근에서 8個地所와 원적암 부근의 9個地所를 택하여 (15×15)m 크기의 quadrat를 각 地所에 5~7개씩 설치하여 出現되는 植物 頻度를 조사하여 78種×52 quadrat의 資料를 얻어 RANK program으로 이를 24種×32 quadrat로 줄여서 分析에 이용하였다(Table 2).

結 果

Sampling information 資料標集에서 information loss가 最小일때의 k (quadrat 數)와 intensity (n)을 정했다. 본 硏究에서 n=25일때로 카나다의 Sifton Bog에서는 0.25 m² (0.5 m×0.5m)의 크기의 quadrat를 이용했고 내장산에서는 225 m²(15m×15m) 크기의 quadrat에서 植物의 頻度를 이용했다.

Fig. 4에서와 같이 Sifton Bog 資料(Table 1)에서는 k=5 일때 information loss가 32%, k=20에서 1%였으며 내장산 자료(Table 2)에서는 k=5 일때 information loss가 32%, k=20 일때는 0.6%였다.

Information loss는 觀側된 要素(f_{ij}, $n-f_{ij}$)와 期待 要素(k, $n-k$)의 차이를 계산한 것으로 $f_{ij} \leqq k$ 일때 식(11)을 이용하여 계산된 것이다. Sifton Bog의 植物 頻度는 n=25가 적절하나 내장산에서는 n=23에서 information loss가 0이 되므로 標集할 때 時間과 經濟的인 面을 考慮하여 quadrat 크기를 225 m²(15m×15m)로 調査하는 것이 효과적인 것 같다.

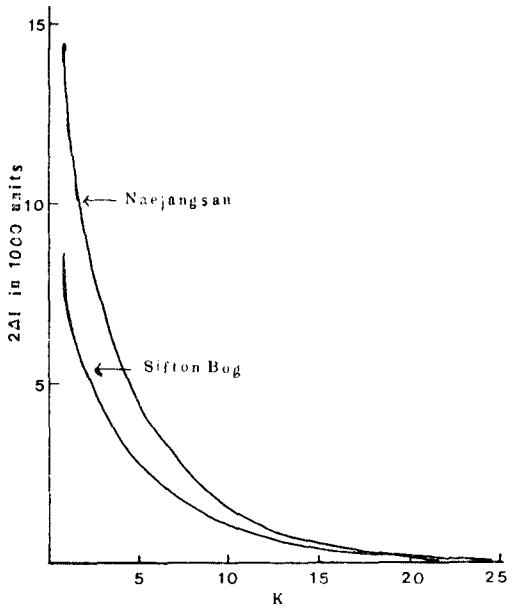

Fig. 4. Information loss curve. The sampling intensity(n) in Sifton Bog data (Table 1) represented 25 and it is 23 in Naejangsan data (Table 2).

Rank에 의한 資料 整理 Table 1은 37種×30 quadrat의 data를 equivocation information으로 種에 대한 rank를 정하여 information rank 값이 17.397 (10%이상)인 20種을 擇하여 20種×30 quadrat로 정리하였고 Table 2는 78種×52 quadrat 크기의 Data를 같은 方法으로 rank에 의해 information rank 값이 16.215 (10%) 이상인 24種을 擇하여 24種×32 quadrat로 정리하였다(Table 3).

R과 Q array와 weighting information 및 CLASS program을 이용하여 Table 1에서 rank가 1과 2인 *Sphagnum fuscum*과 *Sphagnum palustre*에 대하여 class 限界를 0, 1~5, 6~10, 11~15, 15~20, 21~25의 6等級으로 區分하여 R-array하고 marginal dispersion을 구하고 weighting information을 계산하였다(Table 4).

Table 3. Determination of rank for species based on an analysis (RANK) of equivocation information

Species of Sifton Bog	Rank	Information	Species of Naejang San	Rank	Information
Sphagnum fuscum	1	58.4582555	*Lespedeza bicolor*	1	49.6960449
Sphagnum palustre	2	47.2816402	*Lespedeza maximowicz*	2	48.1873107
Vaccinium oxycoccus	3	43.9568073	*Sasa purpurascens*	3	46.3309865
Rhamnus frangula	4	43.7922082	*Fraxinus sieboldiana*	4	45.921967
Rhynchospora alba	5	40.437303	*Pueraria thunbergiana*	5	44.8203105
Chamaedaphne calyculata	6	35.8806137	*Cornus controversa*	6	44.039641
Andromeda glaucophylla	7	31.1129195	*Daphniphyllum macropodum*	7	43.1843711
Drosera rotundifolia	8	30.0778606	*Acer japonica*	8	40.8669782
Sphagnum recurvum	9	24.9302653	*Carpinus laxiflora*	9	38.592689
Hypericum verginianum	10	24.5995935	*Akebia quinata*	10	37.2623754
Thelypteris palustris	11	23.2824387	*Pinus densiflora*	11	36.8732983
Sphagnum riparium	12	23.0932848	*Rhododendron schlippenbachii*	12	36.0165513
Picea mariana	13	23.0657299	*Actinidia arguta*	13	34.295363
Acer rubrum	14	22.9426326	*Staphylea bumalda*	14	29.6390097
Carex sp	15	22.3462548	*Carpinus tschonoskii*	15	26.6522889
Pogonia ophioglssoides	16	21.8723649	*Celastrus orbiculalus*	16	25.983407
Impatiens biflora	17	20.1777315	*Hydrangea serrata*	17	25.1403285
Solanum dulcamara	18	18.7914371	*Zelkova serrata*	18	23.5001815
Sphagnum capillaceum	19	17.7154399	*Styrax japonica*	19	22.4934058
Sarracenia purpurea	20	17.3974551	*Smilax china*	20	22.1807098
			Acer mono	21	22.118169
			Rhus javanica	22	16.8102822
			Benzoin obtusilobum	23	16.370776
			Meliosma myriantha	24	16.2155282

Table 4. R dispersion array for a two species set

A) Raw Data by Table 1.

	quadrat number																												
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1. *S. fuscum*	12	13	10	23	14	0	13	0	13	24	21	17	9	18	19	19	16	23	22	20	8	9	5	17	0	12	0	0	13	21
2. *S. palustre*	1	18	16	4	5	0	8	0	5	5	5	10	14	3	14	4	4	4	9	12	18	17	21	5	0	18	0	0	11	2

B) Marginal dispersion arrays

Class limits	0	1∼5	6∼10	11∼15	16∼20	21∼25
1. *S. fuscum*	5	1	4	7	7	6
2. *S. palustre*	5	12	3	4	5	1

C) R-dispersion array

		species 2						
		0	1∼5	6∼10	11∼15	16∼20	21∼25	Total
species 1	0	5	0	0	0	0	0	5
	1∼ 5	0	0	0	0	0	1	1
	6∼10	0	0	0	1	3	0	4
	11∼15	0	3	1	1	2	0	7
	16∼20	0	4	1	2	0	0	7
	21∼25	0	5	1	0	0	0	6
	Total	5	12	3	4	5	1	30

Table 5. Q dispersion array for a two quadrats set

A) Raw Data by Table 1.

	species number																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. quadrat 11	21	5	12	3	14	18	6	6	5	9	0	0	0	3	0	12	0	0	0	0
2. quadrat 24	17	5	15	1	19	22	7	1	2	0	0	0	3	10	0	2	0	0	5	0

B) Marginal dispersion array

Class limits	0	1∼5	6∼10	11∼15	16∼20	21∼25
1. quadrat 11	8	4	3	3	1	1
2. quadrat 24	7	7	2	1	2	1

C) Q-dispersion array quadrat

		quadrat 24						
		0	1∼5	6∼10	11∼15	6∼20	21∼25	Total
quadrat 11	0	6	2	0	0	0	0	8
	1∼ 5	0	3	1	0	0	0	4
	6∼13	1	1	1	0	0	0	3
	11∼15	0	1	0	1	1	0	3
	16∼20	0	0	0	0	0	1	1
	21∼25	0	0	0	0	1	0	1
	Total	7	7	2	1	2	1	20

Table 6. Results of R and Q dispersion array

A) R-array (Table 4)

1. Information of species 1: I (Sp1) =80.1258
2. Information of species 2: I (Sp2) =75.7142
3. Joint Information; I (Sp1, Sp2) =92.6024
4. Mutual Information: I (Sp1; Sp2) =63.2376
5. Equivocation Information=29.3648
6. Rajski's metrics = 0.3171
7. Coherence coefficient = 0.9484

B) Q-array (Table 5)

1. Information of quadrat 1: I (Q11)=37.7339
2. Information of quadrat 2: I (Q24)=42.1344
3. Joint Information: I (Q11, Q24) =49.1641
4. Mutual Information: I (Q11; Q24)=30.7042
5. Equivocation Information=18.4599
6. Rajski's metrics = 0.3755
7. Coherence coefficient = 0.9268

또한 Q-array는 Table 1에서 quadrat 11과 24에 대한 등급을 R-array와 같이 6等級으로 區分하여 marginal dispersion을 정리하여 weighting information을 계산했다(Table 5). 그 結果는 Table 6와 같다.

Information/Entropy에 의한 Clustering Information/entropy program (INF/ENT)으로 Table 1과 2에서 7가지 information을 선택적으로 계산할 수 있으며 Table 7은 Table 1의 joint information 結果이고 Table 8은 Table 2의 mutual information 結果이다.

또한 clustering은 Table 1과 2에서 equivocation information과 Rajski metrics를 계산하여

Table 7. Result of joint information based on the data from Table 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	80.1	92.6	94.0	98.7	90.7	94.0	89.8	91.2	89.8	87.9	91.8	88.4	92.3	97.4	90.7	86.8	88.2	88.2	85.1	84.3
2		75.7	91.2	97.9	92.6	91.2	87.9	92.6	86.5	86.5	84.3	83.8	92.3	94.6	86.5	86.3	83.8	83.8	82.7	84.6
3			73.0	96.5	83.7	91.2	82.9	89.8	83.8	80.5	86.6	82.3	94.0	93.2	82.3	81.9	85.2	85.2	77.1	82.9
4				62.6	92.3	89.8	88.5	94.6	85.2	78.5	72.9	82.3	90.8	93.7	84.6	73.9	65.1	65.1	78.2	79.7
5					54.0	89.3	67.3	79.6	69.4	68.1	77.9	69.4	81.0	90.4	61.4	69.6	71.6	71.6	67.3	66.7
6						66.8	84.6	88.4	85.7	80.4	83.2	80.1	89.0	94.6	83.8	76.8	74.9	74.9	77.1	77.7
7							43.3	72.1	69.9	60.0	69.6	67.5	73.2	81.6	60.0	63.7	62.3	62.3	51.1	50.0
8								57.7	76.2	71.9	79.1	75.2	87.6	94.6	69.0	74.9	71.0	71.0	69.4	69.2
9									33.4	55.3	58.1	54.9	71.1	81.3	57.9	54.0	52.4	52.4	52.1	56.2
10										30.7	47.3	53.0	72.0	78.2	55.9	43.3	50.6	50.6	51.1	51.2
11											50.7	56.7	74.3	79.6	61.2	44.7	40.7	40.7	48.6	54.5
12												32.6	71.3	79.6	54.9	50.7	52.1	52.1	53.0	53.4
13													51.9	90.4	70.8	70.2	67.0	67.0	63.5	65.9
14														61.1	83.5	75.4	71.6	71.6	72.7	74.1
15															34.7	56.2	54.0	54.0	52.1	51.1
16																27.1	43.3	43.3	44.7	50.7
17																	21.6	21.6	44.7	46.9
18																		21.6	44.7	46.9
19																			24.4	42.6
20																				27.1

Table 8. Result of mutual information based on the data from Table 2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	61.9	25.3	37.3	32.1	18.4	26.1	22.3	28.4	22.3	24.0	28.8	23.9	24.7	16.8	14.2	7.7	16.2	8.3	8.6	16.5	10.1	4.4	8.1	9.8
2		59.9	32.9	24.0	22.5	18.5	23.8	27.7	24.0	25.2	21.4	15.4	19.8	17.8	12.7	4.8	11.3	7.2	8.8	12.8	10.8	5.7	5.8	11.8
3			70.4	29.3	27.8	30.4	25.0	36.4	28.0	29.7	30.3	31.8	30.7	23.0	17.6	14.4	17.2	13.8	16.4	16.7	12.6	5.7	10.0	8.1
4				54.6	17.8	30.3	17.2	25.7	22.5	19.1	25.1	23.8	17.4	13.8	9.6	5.0	10.6	6.8	7.1	9.1	8.0	3.2	4.6	5.6
5					51.9	15.2	14.1	21.2	24.2	23.1	19.3	14.8	21.7	15.7	15.3	8.4	12.3	12.6	11.4	9.9	6.2	9.4	8.6	5.9
6						53.2	15.8	23.5	19.2	13.5	20.1	22.1	20.2	15.8	10.9	6.8	9.7	7.4	7.7	4.9	7.4	3.0	6.3	6.7
7							47.8	17.6	20.4	20.3	12.1	15.6	11.4	17.4	11.2	10.1	8.8	12.5	11.8	10.9	9.9	3.8	5.5	8.2
8								59.7	27.9	27.2	20.7	15.2	26.3	16.2	16.2	6.5	15.4	17.2	10.3	11.5	12.6	10.6	11.9	8.1
9									53.0	24.2	18.0	17.4	19.7	11.7	9.7	7.4	11.2	10.7	7.8	10.3	11.2	5.1	4.6	3.1
10										53.0	20.6	13.7	21.0	13.1	10.5	5.5	12.5	13.1	7.8	12.3	6.4	4.2	5.9	5.5
11											48.7	17.7	18.0	11.3	7.6	6.4	11.9	5.3	5.4	13.1	12.5	6.2	3.6	4.5
12												42.7	19.7	13.4	4.5	8.1	7.2	4.2	12.1	9.7	9.1	3.8	2.6	2.0
13													52.1	20.9	12.1	6.1	13.1	9.9	10.3	8.3	8.3	4.8	5.8	6.5
14														42.5	11.5	10.7	9.2	11.6	11.5	9.3	7.9	4.0	5.3	7.3
15															31.9	4.5	15.5	8.0	7.6	8.9	7.1	6.0	13.6	1.2
16																26.0	3.7	3.7	13.3	3.2	1.6	2.2	1.7	2.7
17																	38.1	7.0	7.8	12.4	4.3	4.0	8.6	2.3
18																		25.7	5.2	4.3	4.8	1.3	4.8	3.7
19																			27.7	4.1	3.7	3.5	3.2	3.0
20																				38.9	2.9	4.5	6.0	5.6
21																					35.9	2.6	3.1	7.8
22																						19.7	4.0	2.4
23																							19.1	3.5
24																								17.6

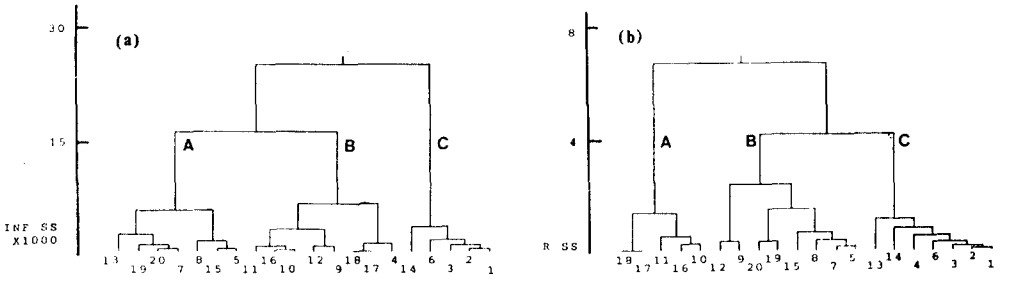

Fig. 5. Dendrograms of 20 species in Table 1. The vertical scale indicates sum of square(SS) of information. The (a) classified by basis on equivocation information and (b) by Rajski's metrics.

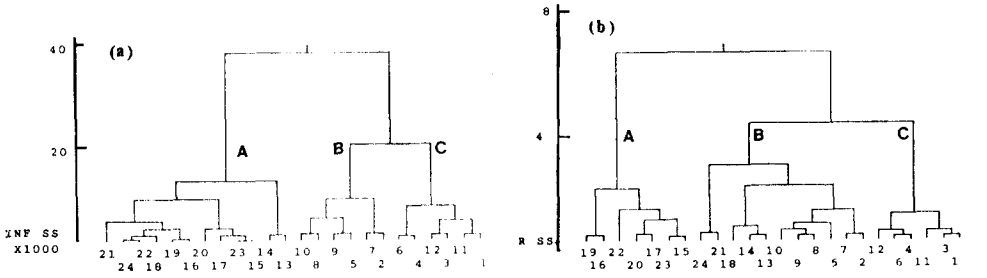

Fig. 6. Dendrograms of 24 species in Table 2. The vertical scale indicates sum of square of information. The (a) classified by basis on equivocation information and (b) on Rajski's metrics.

sum of square(SS) algorithm으로 fusion해서 Fig. 5(a, b)와 Fig. 6(a, b) 같이 dendrogram으로 나타냈다.

Fig. 5의 clustering은 Sifton Bog 資料(Table 1)의 結果로 Fig. 5(a)는 equivocation information 값으로 cluster해서 SS-information이 15×1,000 일때 區分했고 Fig. 5(b)는 Rajski metrics 값을 이용한 것으로 SS-information이 4일때 區分한 것이다. 이때 cluster의 效率은 equivocation information으로 clustering 했을때는 全體 cluster SS가 30248, cluster內의 SS가 18695이므로 cluster間의 SS는 11553이 된다. 그래서 效率은 38.2% (11553/30248×100)가 된다. 반면에 Rajski metrics에 의한 clustering에서는 전체 cluster SS가 5.789, cluster 內의 SS는 4.219로써 cluster間 SS가 1.571로 效率은 27.1%였다.

또한 equivocation information에 의한 clustering에서 cluster A에는 13, 19, 20, 7, 8, 15, 5의 7種이 포함되었고, cluster B에는 11, 16, 10, 12, 9, 18, 17, 4의 8種이 포함되고, cluster C에는 14, 6, 3, 1, 2의 5種이 포함되었다(Fig. 5(a)).

Rajski metric에 의한 clustering 結果는 cluster A는 18, 17, 11, 16, 10의 5種이 포함되며, cluster B에는 12, 9, 20, 19, 15, 8, 7, 5, 의 8種이 포함되었고, cluster C에는 15, 14, 4, 6, 3, 2, 1의 7種이 포함 되었다(Fig. 5(b)).

두 方法에 의한 clustering의 결과에서 cluster C는 거의 비슷하나 cluster A와 B는 큰 차이가 있었다.

Fig. 6는 內藏山의 資料(Table 2)의 clustering의 結果로 Fig. 6(a)는 equivocation information에 의한 것이며, Fig. 6(b)는 Rajski metrics에 의한 것이다. equivocation information에 의한 clustering 效率은 全體 cluster의 SS가 43882이고, cluster內의 SS가 31989이므로

cluster間의 SS는 11894로서 效率은 27.1%(11894/43882×100)였으며 Rajski metrics에 의한 clustering 效率은 全體 cluster의 SS가 8.144이고, cluster內의 SS가 6.721이므로 cluster間의 SS가 1.421이므로 區分 效率이 17.4%(1.421/8.144×100)으로 나타 났다.

역시 내장산의 資料에 대한 cluster 區分도 3개의 cluster로서 equivocation information에 의한 clustering에서 cluster A는 21, 24, 22, 18, 19, 16, 20, 17, 23, 15, 14, 13으로 12種이 포함되며, cluster B는 10, 8, 9, 5, 7, 2으로 6種이 포함되며, cluster C에는 6, 4, 12, 3, 11, 1의 6種이 포함되었다. 반면에 Rajski metrics에 의한 clustering에서는 cluster A에는 19, 16, 22, 20, 17, 23, 15의 7種이 포함되고 cluster B에는 24, 21, 18, 14, 13, 10, 9, 8, 5, 7, 2의 11種이 포함하고 cluster C에는 12, 6, 4, 11, 3, 1의 6種이 포함 되었다.

두 方法에 의한 clustering의 結果에서도 cluster C에서는 비슷한 양상을 띠었으나 cluster A와 B는 큰 차이가 있었다. 본 硏究에서는 clustering의 效率이 높은 equivocation information에 의한 clustering의 結果를 이용하여 集中度分析을 하였다.

集中度分析(AOC) 集中度分析을 하기 위하여 分割表는 equivocation information에 의한 cluster A(12種), B(6種) C(6種)과 北斜面 Ⅰ地所(전망대 ; 6 quadrat), Ⅱ地所(선인봉 ; 9 quadrat)와 南斜面에서 Ⅲ地所(까치봉 ; 8 quadrat), Ⅳ地所(원적암 ; 9 quadrat)의 3×4 Block을 정리하여 식(18)을 이용하여 coherence coefficience를 種(24種)에 대하여 算出된 값(I_{pp})

Table 9. Sum of Rajski's metrics and blocks among three clusters and four stands, and adjusted table correspondeding with block table and information

	Sum of Rajski s metrics				Block size				Adjusted value			
Stands	I	II	III	IV	I	II	III	IV	I	II	III	IV
Cluster A	43.16	58.99	52.26	62.51	72	108	96	108	36.52	33.28	33.17	35.26
Cluster B	27.34	37.59	34.20	43.10	36	54	48	54	46.27	42.41	43.41	48.63
Cluster C	28.48	37.36	37.15	44.47	36	54	48	54	48.20	42.15	47.15	50.17

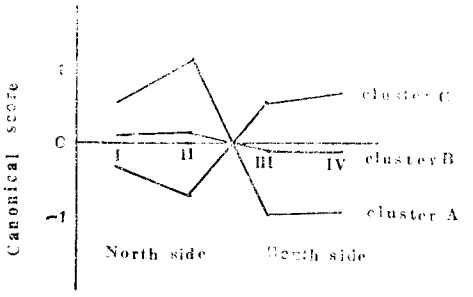

Fig. 7. Expected deviation resulting analysis of concentration of three clusters and four stands in Naejangsan data. Cluster A reveals the different gradient along north and south sides against cluster B.

과 Table 2 (R_{pq})를 이용하여 block에 따라 合算한 후 adjust했다(Table 9).

集中度分析 次數의 결정은 Gittins(1979)의 理論에 따라 第一變量은 79%, 第二變量은 21%의 分析力을 나타냈는데 第一變量만을 이용하여 해석했다.

Fig. 7에 나타난 바와 같이 cluster A에 속하는 고추나무, 개서나무, 산수국, 노박덩굴, 때죽나무, 고로쇠, 생강나무등은 북사면에 높게 나타났고, 반대로 cluster C에 속하는 싸리나무, 조릿대, 쇠물푸레나무, 소나무, 층층나무 등은 남사면에 높게 나타났으나 cluster B에 속하는 굴거리나무, 참

단풍, 서나무, 칡등은 남북사면에 차이 없이 나타났다.

論 議

意思傳達理論(communication theory)에 의한 information理論의 適用은 새로운 理論은 아니나 여러 學問에 多樣하게 適用되어 왔다. 生物學에 이 理論을 적용한 것은 細胞生物學 部分에 Yockey(1958) 等이 이용한 것이 처음이며, 種多樣性의 定量에 Shannon指數를 Magalef(1958)가 적용한 후 Pielou(1966), Orloci(1977)등에 의해서 information 理論을 生態學에 實用 됐으며 Feoli *et al.* (1984)은 FORTRAN program을 작성하였고 Orloci(1978)는 BASIC program으로 data 分析을 할 수 있게 하였다.

본 硏究에서는 Fig. 2와 같이 標集할 때 information loss가 最小일 때의 sampling intensity를 정하여 quadrat數와 크기를 限定하여 시간과 노력을 최소화할 수 있고 標集의 效果를 산출할 수 있게 하였다.

Table 1의 資料에서 n=25일 때 information loss가 k=5에서 32%였고 k=25에서 0이 됐다. 이때 quadrat의 크기는 (0.5×0.5)m로 最適의 크기였다. 또한 Table 2에서는 n(intensity)이 23일때 k=5에서 information loss가 32%였으나 k=23에서 0이됐고 quadrat크기가 (15×15)m로 종래의 森林生態學에서 (10×10)m의 크기의 quadrat보다 내장산에서의 이 연구에서는 약간 클때 標集效率이 높은 것으로 나타났다.

Rank를 정할때 Orloci(1978)는 equivocation information과 mutual information을 대비했는데 效率이 좋은 equivocation information을 이용하였고 class의 限界는 임의로 택하였다.

計算된 information/entropy 값으로 clustering의 區分은 agglomerative algorithm 法으로 (Sneath and Sokal, 1973) sum of square algorithm을 이용 했으며, 이때 주로 equivocation information과 Rajski metrics를 이용했다. Table 1의 clustering에서 equivocation information으로 구분하면 效率이 38.2% 였으나 Rajski metrics에서는 27.1% 였으며, Table 2의 clustering 效率은 前者에 의해서 37.2%, 後者에 의해서 17.4%로 나타났다. 따라서 集中度分析은 equivocation information clustering 結果를 이용하여 분석했다.

같은 植物群落內에서 植物과 環境과의 關係는 類型(pattern)에 따라 環境變數에 의한 勾配分析(gradient analysis; Whittaker, 1967)과 어떤 地域內의 aggregation된 植物과 環境과의 關係를 ordination 또는 canonical analysis (Orloci, 1978; Gauch, 1982)를 이용하여 classification해서 Block間의 관계를 AOC法으로 解析하는 것이 보통이다(Feoli and Orloci, 1979; Park, 1984).

內藏山 資料(Table 2)에 대한 集中度分析(AOC)을 위해서 coherence coefficient를 산출하여 (I_{pp}) adjust는 3×4 block을 clustering으로 정하고 $I_{pp}R_{pq}/N$(식(19))로 변형하여 block 內의 合을 이용하였다(Table 9). 3個의 cluster(A, B 및 C)와 北斜面(Ⅰ, Ⅱ地所)과 南斜面(Ⅲ, Ⅳ地所)에 대한 集中度分析의 結果는 第一變量이 79%, 第二變量이 21%로 나타나 第一變量으로만 해석했다.

Fig. 7과 같이 cluster A에 포함되는 植物은 北斜面에 높은 頻度를 나타내고 cluster C에 포함되는 植物은 南斜面에 높게 나타났고 cluster B에 포함되는 植物은 南北斜面에 차이가 없이 나타났다.

Information/entropy에 의한 結果 해석은 종래의 多樣性이나 clustering으로 表現해서 資料를 해석하는 이외에도 集中度分析으로 資料內의 關係를 구체적으로 해석하는데 이용할

수 있다.

摘 要

Information과 entropy理論을 適用하여 카나다 Sifton Bog 資料(Table 1)와 內藏山에서 조사된 資料(Table 2)를 分析하였다.

標集할때 information loss가 最小일 때의 sampling intensity (n)를 限定할 수 있게 했고, weighting information을 算出할때 class를 임의로 區分하여 information 값을 計算했으며, 資料가 방대할 때 equivocation information으로 rank를 정하여 資料의 크기를 줄일 수 있게 했다.

또한 information/entropy 計算을 7가지로 구분하여 선택적으로 information을 算出할 수 있게 하였다. information에 의한 結果를 解析할 때 多樣性 계산과는 다르게 entropy 값으로 clustering하고 效率을 計算하였다.

植被 資料에서 種의 cluster와 調査地所와의 關係 分析은 集中度法을 이용하여 內藏山 資料에서 3個의 種 cluster와 4個 調査地所에 대하여 Rajski metrics를 adjust해서 分析하였다. 이때 第一變量(79%) 만을 이용하여 解析하였다.

Cluster A에 속하는 개서나무, 노박덩굴, 고로쇠, 생강나무, 때죽나무등은 北斜面에 높게 나타났으며 cluster C에 속하는 싸리나무, 소나무, 쇠물푸레나무, 층층나무등은 南斜面에 높게 나타났으나 cluster B에 속한 참단풍, 서나무등은 남북사면에 관계없이 고르게 나타났다.

引 用 文 獻

Feoli, E. and L. Orloci. (1979). Analysis of concentration and detection of underlying factors in structure tables. Vegetatio, **40** : 49~54.

Feoli, E., M. Lagonegro and L. Orloci. (1984). Information analysis of vegetation data. Junk, Hague.

Fisher, R.A. (1948). Statistical method for research workers. 10th ed. Oliver and Bayd., Edinburgh.

Gauch, H.G. Jr. (1982). Multivariate analysis in community ecology. Cambridge Univ. Press, Cambridge.

Greig-Smith, P. (1957). Quantitative plant ecology. Butterworth, London.

Kullback, S. (1959). Information theory and statistics. Wiley, New York.

MacArthur, R.H. (1965). Pattern of species diversity. Biol. Rev., **40** : 510~533.

Margalef, D.R. (1958). Information theory in ecology. Yearbook Soc. Syst. Research, **3** : 36~71.

McIntosh, R.P. (1967). An index of divesity and the relation of certain concepts to diversity. Ecology, **48** : 392~403.

Legendre, L. and P. Legendre. (1983). Numerical ecology. Elsevier Sci. Co., New York.

Orloci, L. (1970). Analysis of vegetation samples based on the use of information. J. Theor. Biol., **29** : 173~189.

Orloci, L. (1972). On information analysis in phytosociology. Junk, Hague.

Orloci, L. (1976). Ranking species by an information criterion. J. Ecol., **64** : 417~419.

Orloci, L. (1977). Ranking species based on the components of equivocation information. Vegetatio, **37** : 123~125.

Orloci, L. (1978). Multivariate analysis in vegetation research. Junk, Hague.

Park, B.K. (1974). A phytosociological study on the vegetation of National Park Mt. Naejangsan. A report of Korea. Associ. for Conservation of Nature.

Park, S.T. (1984). An application of analysis of concentration for ecological study. Korean J. Bot., **27** : 223~231.

Pielou, E.C. (1966). Shannon's formula as measure of species diversity. Amer. Natur., **100** : 463~465.

Pielou, E.C. (1977). Mathematical ecology. Wiely, New York.

Rajski, C. (1961). Entropy and metric space: Information theory. Butteworth, London.

Renyi, A. (1961). On measures of entropy and information. 4th Berkeley Symposium on Math. Statis. and Prob. Univ. of Calif. Press, Berkeley.

Shannon, C.E. (1948). A mathematical theory of communication. Bell system Tech., **27** : 379~423.

Sneath, P.H.A. and V.D. Sokal. (1973). Numerical taxonomy. Freeman, San Francisco.

Whittacker, R.H. (1967). Gradient analysis of vegetation. Biol. Rev., **42** : 207~264.

Yockey, H.P., R.L. Platzman and H. Quastler. (1958). Information theory in biology. Pergamon, New York.

((1987年 3月 23日 接受)