数値実験の比較による、関係類似度を用いた Rsn 法の利点

品川 汐夫, 多部田 修

(1997 年 3 月 3 日受付)

Advantages of Rsn Method over Correspondence
Analysis Used for Simulated DataSekio Shinagawa*¹ and Osame Tabeta*²

In order to indicate the advantages of the Rsn method, we compared it with the widely used method, correspondence analysis (CA) based on the data simulated to those of benthic macrofauna. It is proved that the Rsn method has the following advantages. The sites of similar species composition are displayed in a mass, and their patterns correspond well with the spatial distributions of the species, that is preferable to the cluster analysis. The cumulative ratio of eigen values of the Rsn matrix is relatively high. This means that an error component included in the matrix is low, while the essential information is well preserved. The Rsn method, therefore, is applicable to the data analysis of benthic macrofauna.

キーワード: 群集解析, 座標づけ, 類似度指数, 多変量解析, 関係類似度, 対応分析

種×地点の2次元のデータに基づき、生物群集の構造を客観的に把握するための数理的方法として、座標づけとクラスター解析がある。^{1,2)} 一般に、座標づけは比較的単純なデータ構造の解析に、クラスター解析はより複雑かつノイズの多いデータ構造の解析に適した方法とされる。³⁾ 座標づけはまた、適切なクラスタリング結果を得るための前処理としても用いることができる。座標づけに対応分析 (CA) を用いた、TWINSPAN^{4,5)} によるクラスタリングはその例である。この CA は、我が国では数量化Ⅲ類とも呼ばれる極めて応用範囲の広い方法で、1970 年代以降パソコンの普及とともに、群集の数理解析の分野でも盛んに用いられるようになってきた。最近では、CA を改良した除歪法 (DCA)⁶⁾ が高く評価され、とくに植生学の分野での適用例が多い。しかし、DCA のプログラムの詳細なアルゴリズムは必ずしも明らかではないが、⁶⁾ 基本的な点においては、DCA と CA との間には大きな相違はないように思われる。また加藤⁷⁾ は、DCA にも問題点があることを指摘して、CA と併用することが望ましいとしている。

一方、海域のマクロベントス群集の解析においては、これらの方法の適用例はまだ少ない。⁸⁾ その理由の一つとして、マクロベントスの採集データは、少数の優

占種とともに、偶然性によって出現が左右される個体数が少ない種を多数含み、かつ採集上の問題などから不規則な変動が大きいこと、CA による解析が必ずしも好ましい結果を与えないことにあると思われる。このような変動を抑えるためにはサンプリング面積を増やしたり、採集方法を改良することなどが考えられるが、実際のモニタリング調査などでは費用などの点で限界がある。石川⁹⁾ は、マクロベントスと底質の調査データに正準相関分析を適用して、マクロベントスのデータには誤差が多く含まれているため、誤差の少ない、いくつかの優占種を選んで解析すべきこと、および底質データの解析を先ず行うべきであることを述べている。しかし生物群集には、たとえ個体数は少なくとも理化学的データからは得られない情報をもつ種も存在するから、そのような種も含めた、誤差の多いデータから意味のある情報を引き出し、群集自身に語らせる²⁾ 解析方法の確立が望まれる。

品川¹⁰⁾ は、このようなマクロベントスの採集データの解析法として、地点間相関指数 Rs と関係類似度 Rsn を定義し、Rsn を成分とする行列の固有値分解による座標づけと、その結果に基づくクラスター解析 (Rsn 法) を提案した (座標づけ方法の概略は本論文の末尾に付録として示した)。また、この方法を底生動物調査や

*¹ 下関女子短期大学 (Shimonoseki Women's Junior College, Shimonoseki 750, Japan).

*² 長崎大学水産学部 (Faculty of Fisheries, Nagasaki University, Nagasaki 852, Japan).

*³ 品川汐夫: 魚族相の解析, 下関市南部水域資源診断事業に関する報告書, 下関外海漁業共励会, 下関, 1983, pp. 27-43.

漁獲物組成, さらに魚類の胃内容物組成の解析に適用し, Rsn 法は, ノイズの多いデータにおいても, その中に潜む類似したグループをよく検出し, そこから意味のある情報を引き出す上で極めて有効であるという結果を得ている。^{11),*3}

そこで著者らは, この実際のデータで得られた Rsn 法の利点をより客観的に示すことと, CA との相違点とその問題点を検討することを目的として, マクロベントス群集に模した人工データを作成して数値実験を行な

い, 解析上問題となる特徴的な群集構造が, CA と Rsn 法ではどのように表示されるか, さらに不規則な変動が加わった場合の解析の安定性などについて比較研究を行った。

材料および方法

非確率の人工データによる数値実験 実験に用いたデータは, Table 1 に示す CASE-1~4 の 4 通り, それぞれ 14 種×10 区画である。CASE-1 は, 動物相が環境

Table 1. Four cases of the simple, simulated data matrix, each with 14 rows (species) and 10 columns (quadrats)

CASE-1										
	1	2	3	4	5	6	7	8	9	10
SP-1	1	0	0	0	0	0	0	0	0	0
SP-2	2	1	0	0	0	0	0	0	0	0
SP-3	3	2	1	0	0	0	0	0	0	0
SP-4	2	3	2	1	0	0	0	0	0	0
SP-5	1	2	3	2	1	0	0	0	0	0
SP-6	0	1	2	3	2	1	0	0	0	0
SP-7	0	0	1	2	3	2	1	0	0	0
SP-8	0	0	0	1	2	3	2	1	0	0
SP-9	0	0	0	0	1	2	3	2	1	0
SP-10	0	0	0	0	0	1	2	3	2	1
SP-11	0	0	0	0	0	0	1	2	3	2
SP-12	0	0	0	0	0	0	0	1	2	3
SP-13	0	0	0	0	0	0	0	0	1	2
SP-14	0	0	0	0	0	0	0	0	0	1

CASE-2										
	1	2	3	4	5	6	7	8	9	10
SP-1	14	13	12	11	10	9	8	7	6	5
SP-2	13	12	11	10	9	8	7	6	5	4
SP-3	12	11	10	9	8	7	6	5	4	3
SP-4	11	10	9	8	7	6	5	4	3	2
SP-5	10	9	8	7	6	5	4	3	2	1
SP-6	9	8	7	6	5	4	3	2	1	0
SP-7	8	7	6	5	4	3	2	1	0	0
SP-8	7	6	5	4	3	2	1	0	0	0
SP-9	6	5	4	3	2	1	0	0	0	0
SP-10	5	4	3	2	1	0	0	0	0	0
SP-11	4	3	2	1	0	0	0	0	0	0
SP-12	3	2	1	0	0	0	0	0	0	0
SP-13	2	1	0	0	0	0	0	0	0	0
SP-14	1	0	0	0	0	0	0	0	0	0

CASE-3										
	1	2	3	4	5	6	7	8	9	10
SP-1	14	13	12	11	10	9	8	7	6	5
SP-2	13	12	11	10	9	8	7	6	5	4
SP-3	12	11	10	9	8	7	6	5	4	3
SP-4	11	10	9	8	7	6	5	4	3	2
SP-5	10	9	8	7	6	5	4	3	2	1
SP-6	9	8	7	6	5	4	3	2	1	0
SP-7	8	7	6	5	4	3	2	1	0	0
SP-8	0	0	1	2	3	4	5	6	7	8
SP-9	0	1	2	3	4	5	6	7	8	9
SP-10	1	2	3	4	5	6	7	8	9	10
SP-11	2	3	4	5	6	7	8	9	10	11
SP-12	3	4	5	6	7	8	9	10	11	12
SP-13	4	5	6	7	8	9	10	11	12	13
SP-14	5	6	7	8	9	10	11	12	13	14

CASE-4										
	1	2	3	4	5	6	7	8	9	10
SP-1	14	13	12	11	10	9	8	7	6	5
SP-2	13	12	11	10	9	8	7	6	5	4
SP-3	12	11	10	9	8	7	6	5	4	3
SP-4	11	10	9	8	7	6	5	4	3	2
SP-5	10	9	8	7	6	5	4	3	2	1
SP-6	5	6	7	8	9	10	11	12	13	14
SP-7	4	5	6	7	8	9	10	11	12	13
SP-8	3	4	5	6	7	8	9	10	11	12
SP-9	2	3	4	5	6	7	8	9	10	11
SP-10	1	2	3	4	5	6	7	8	9	10
SP-11	4	5	6	7	8	8	7	6	5	4
SP-12	3	4	5	6	7	7	6	5	4	3
SP-13	2	3	4	5	6	6	5	4	3	2
SP-14	1	2	3	4	5	5	4	3	2	1

CASE-1: One dimensional data structure, each adjacent quadrat having the same difference. CASE-2: All species have the same distribution pattern but different abundance. CASE-3: Two symmetrically distributed species groups are present, each having its climax at the either end. CASE-4: In addition to two species groups similar to CASE-3, one fewer species group having its climax at the center is present.

勾配に沿って連続的に変化する場合で、低所～高所での植物群落の遷移などにみられる構造である。これは、クラスター解析には不向きな構造が、CA と Rsn 法による座標づけではどのように示されるかをみるためのものである。またこれは、多くの座標づけ方法の欠点とされるアーチ効果を示す典型として、よく取り上げられている。¹⁾ CASE-2 は、全種が同じ単調な変化をした場合である。種組成に大きな変化はなく、総個体数が変化する地点間の相違が、座標づけではどのように示されるかをみるためのものである。実際のベントス調査においても、採集個体数が少ないことがその地点の特徴になる場合がある。サンプルサイズに影響されないとされる Morishita の類似度 C_d などではその相違は検出し難い。CASE-3 は、二つの相反する単調変化がある場合で、河口域干潟など、環境勾配に沿って種組成が大きく変化する場合を想定している。CASE-4 は、CASE-3 にさらに個体数の小さい山型変化が加わった場合である。このような環境勾配に沿った山型変化の存在が、単調変化を前提とした一般的多変量解析の群集解析への適用を困難にしている。ここではそのような構造が、CA と Rsn 法による座標づけではどのように示されるかをみることを目的とした。

確率的人工データによる数値実験 データの不規則な変動が、CA と Rsn 法による解析結果におよぼす影響を調べるため、乱数を用いて確率的変動を含む人工データをつぎのように作成し、計算結果を比較した。

x 方向 7×y 方向 4 の正方形で区分された長方形の領域 (Fig. 1) で、個体の位置を、x 方向は正規分布 ($-\infty \sim +\infty$) に、y 方向は 1~4 間の一様分布に従う確率変数を用いて定め、種別に各区画内の個体数を合計して、種×区画のデータを、CASE-5~7 の 3 通り作成した。作成したデータの、個体の位置の正規分布パターンは Fig. 2 に示す A~G の 7 通りとした。各 CASE について、A~G の種類数と 1 種当たりの総個体数および、

Fig. 1. Spatial arrangement of quadrats used to generate the stochastic, simulated data.

Symbols in each quadrat denote the group of quadrats located along the x-axis. They are used in the left graphs of Fig. 5 and 6.

Fig. 2. The distribution pattern of each species group A~G along the x-axis used to generate the stochastic, simulated data in each case.

Each curve represents a part of the normal probability density (P) of the location of an individual ranging from $-\infty$ to $+\infty$ along the x-axis.

正規分布の平均値と標準偏差を Table 2 に示す。表に示すように、各 CASE ともマクロベントス群集のデータに模するため、出現が偶然性によって左右される個体数が少ない種を多く含み、個体数が多い種を少数含むように配慮した。CASE-5 は、Rsn の計算で問題となる、全種が平均的に出現する中央部の区画の配置を調べるためであり、CASE-6 は、著しい優占種 (E) の影響と、大きな種組成の変化がなく総個体数が変化する、中央部より右側の区画の配置を調べるためである。また CASE-7 は、広範囲に出現する著しい優占種 (E) と、中央部を中心に出現する少数種 (G) が、ともに存在する場合を想定している。ベントス調査では、特定の地点に特定の少数種が出現し、それが環境を考える上で重要な示唆を与えることがある。例えば沖合水の影響を示唆する特定の端脚類が少数出現する場合などである。個体数の多い種の影響が大きい類似度による解析では、そのような地点の特徴を示すことができない。CASE-7 は、CA と Rsn 法による座標づけではそれがどのように示されるかをみることを目的としている。Fig. 1 には各区画の位置と、以下で用いるそれらの記号を示した。

結 果

非確率的人工データによる数値実験

Table 2. The number of species and the total number of individuals per species with the assumed mean and standard deviation (S.D.) of the location of an individual along the x-axis (see Fig. 2.), for generating the stochastic, simulated data with normal distributed random variables

CASE-5					CASE-6				
Sp. group	Sp. no.	Ind. no.	Mean	S.D.	Sp. group	Sp. no.	Ind. no.	Mean	S.D.
A	16	16	0	3	C	16	16	0	2
	8	32	0	3		8	32	0	2
	4	64	0	3		4	64	0	2
	2	128	0	3		2	128	0	2
	1	256	0	3	D	16	16	3.5	2
	1	512	0	3		8	32	3.5	2
B	16	16	7	3		4	64	3.5	2
	8	32	7	3		2	128	3.5	2
	4	64	7	3	E	2	1024	3.5	4
	2	128	7	3					
	1	256	7	3					
	1	512	7	3					
CASE-7									
Sp. group	Sp. no.	Ind. no.	Mean	S.D.					
C	16	16	0	2					
	8	32	0	2					
	4	64	0	2					
	2	128	0	2					
F	16	16	7	2					
	8	32	7	2					
	4	64	7	2					
	2	128	7	2					
E	2	1024	3.5	4					
G	3	32	3.5	1					

CASE-5: Two species groups (A) & (B) distributed symmetrically are present, each having its climax at either end. CASE-6: Three species groups, one with the left end climax group (C), the second with the central climax group (D), and the third dominant species group with wide distribution (E) are present. CASE-7: In addition to both species groups (C and F) similar to CASE-5 and one dominant species group (E), one fewer species group (G) having its climax at the center is present.

1) 対応分析 (CA) の結果 第Ⅰ～第Ⅱ座標平面に区画と種を配置した結果を Fig. 3 に示す。各 CASE の固有値の、第Ⅱ成分までの累積率 (全固有値の合計に対する累積和の比率) は、それぞれ 73.0%, 93.0%, 99.9%, 100.0% であった。

CASE-1: よく知られているように、区画と種はいずれも第Ⅰ～第Ⅱ座標平面では 2 次曲線上に、第Ⅰ～第Ⅲ座標平面では 3 次曲線上に配置される。¹²⁾ このため、図示はしていないが、第Ⅲ成分の値も $-0.8 \sim 0.8$ と、とくに端部では大きな値であった。このことは、有意な成分数の取り方によって区画間の関係が異なって解釈されることを示すものである。

CASE-2: 区画と種はいずれも CASE-1 と同様に、ほぼ 2 次曲線上に配置されるが、左右非対称となり、区画と種の配置から、第Ⅰ成分が負の区画には、正の区画よりも多くの種が出現していることなどが読み取れる。

CASE-3: 区画と種はいずれも直線上に配置され、区画はほぼ等間隔であるが、種は二つのグループに顕著に分けられる。

CASE-4: 区画は中央で折れ曲がった直線上に配置され、中央部は端部より間隔が小さくなる。これは、中央部に多く出現する SP-11～14 の影響を表している。一方、種は三つのグループが顕著に区別され、直線上に配

Fig. 3. CA ordinations of 10 quadrats (left) and 14 species (right) onto the axis I $\times$ II plane for the simple, simulated data (Table 1).

Each point is labeled with a number denoting the column (quadrat) and the row (species) of the data matrix, respectively.

置される。

2) Rsn 法の結果 区画と種を, CASE-1~3 については第 I ~ 第 II 座標平面上に, CASE-4 については第 III 成分まで意味のある配置がみられたので, 第 I ~ 第 II および第 I ~ 第 III 座標平面上に配置した結果を Fig. 4 に示す。各 CASE の固有値の, 第 II 成分までの累積率は, それぞれ 90.5%, 89.6%, 93.3%, 82.0% であり, また, CASE-4 の第 III 成分までの累積率は 96.6% であった。Rsn 法では, n 個の区画は n 次元の単位球面上に配置され, 区画間の関係は距離ではなく区画の位置ベクトル間の余弦によって表される。これは主成分分析の因子軸への項目の配置と同じであるが, 区画間の距離によって相違を表す CA とは異なるから, 解釈には注意を要する。図は, それらの 2 次元平面への投影である。また, 区画の原点からの距離は, この 2 次元平面によって説明される割合を表す。

CASE-1: 両端と中央部の区画が, それぞれほぼ正三角形の頂点に配置されており, これらが, それぞれ対等な関係にあることを示している。両端部および中央部で区画間の距離が他より小さくなっており, それぞれ三つの群として認識できる。種についても同様に, 両端と中

Fig. 4. Rsn ordinations of 10 quadrats (left) and 14 species (right) for the simple, simulated data (Table 1).

In CASE-1~3, the axis I $\times$ II plane is shown. In CASE-3, the axis I $\times$ III plane is also shown. Each point denotes the same as in Fig. 1.

央部の三つの群が識別される。第 III 成分は図示はしていないが, 全区画に共通して 0.15~0.40 と小さく, 区画間関係には影響しない程度であった。

CASE-2: 区画は第 I と第 IV 象限に, 種は全て第 IV 象限に配置されており, CASE-1 とは異なったパターンを示す。このことから, 区画間関係には共通の成分 I が大きく働いていること, また種の分布パターンに大きな相違はなく, 区画 1~5 に多くの種が出現していること, さらに SP-1~7 は, 共通の成分 I が大きいことから, 出現範囲が広いことなどが読み取れる。

CASE-3: 区画 5 と 6 の位置ベクトルがほぼ直交している。これは, 区画 5 と 6 の間の関係類似度が 0 に近いので, このためにこの区画を境として左右の区画が大きく隔てられる。これは区画 5 と 6 の相違を誇張する欠点もあるが, 誤差の多いデータから, 異なったグループを検出する場合には利点ともなる。第 II 成分は全区

画に共通する成分である。どの種も第Ⅱ成分が大きいことは、いずれも出現範囲が広いことと対応している。

CASE-4: 第Ⅰ成分によって、区画1~3と8~10の2つの群AとBが分けられ、第Ⅱと第Ⅲ成分によって区画5と6の群Cが分けられている。区画4と7は、それぞれAとCおよびBとCの中間に配置されている。このようにRsn法では、類似した区画が群としてまとめられることが、CAとは異なっている。種の配置では、第Ⅰと第Ⅱ成分によって3つの群が明瞭に区別されるが、いずれも第Ⅱ成分の値が大きいことはCASE-3の場合と同様である。中央部を中心に出現するSP-11~14は、第Ⅲ成分によって、CAの場合と同じ順に内側から配置されている。

確率的人工データによる数値実験

1) 対応分析 (CA) の結果 各CASEの固有値の、第Ⅱ成分までの累積率は、それぞれ37.2%, 35.0%, 44.5%であった。岩坪¹²⁾は、CAの固有値の累積寄与率として、上に述べた累積率の平方根を用いるべきことを示しているが、それでも約60%と低い。これは、無意味な誤差成分の割合が大きいことを意味している。各CASEの区画と種の配置をFig. 5に示す。CASE-5と6では、第Ⅱ成分ですでに不規則な配置がみられ、とくに第Ⅰ軸の端部でそれは大きい。第Ⅲ成分ではさらに不規則な配置であった。またCASE-6では、中央部より右側の区画が、期待される左から右の順に順序づけられ

Fig. 5. CA ordinations of quadrats (left) and species (right) onto the axis I $\times$ II plane for the stochastic, simulated data (Table 2).

Symbols in the left graphs denote the group of quadrats shown in Fig. 1, and those in the right graphs are labeled with a letter denoting the group of species given in Fig. 2.

ていない。この理由は、CAではデータ行列における行和と列和の積の平方根に対するデータの比率が解析の対象となるため、この比率が顕著に変化しない、中央部より右の領域の区画相互間の差異が表せないためと思われる。CASE-7では、著しい優占種 (E) の存在にも関わらずCASE-1または4に類似した配置となり、全ての区画がほぼ正しく順序づけられている。しかし種の配置には飛び離れた点がみられ、群のまとまりが悪い。これはすべてのCASEでみられ、誤差の影響が直接表れていることを示している。

2) Rsn法の結果 各CASEの固有値の、第Ⅱ成分までの累積率は、それぞれ78.5%, 78.7%, 87.6%と、CAの場合に比べて大きい。このことは、無意味な成分の割合がCAの場合より小さいことを意味している。各CASEの区画と種の配置をFig. 6に示す。CASE-5と7では、区画は第Ⅰ~第Ⅱ座標平面上でほぼ単位円周上に配置されており、有意な成分として第Ⅱ成分まで考えれば十分と思われるが、CASE-6では、とくに中央部より右側の区画で円周からのずれが大きく、第Ⅱ成分までは十分な情報が表されていないことを示している。そこでCASE-6については、第Ⅰ~第Ⅲ座標平面上の配置も示した。その結果、CASE-6の中央部より右側の

Fig. 6. Rsn ordinations of quadrats (left) and species (right) for the stochastic, simulated data.

In CASE-5 and 7, the axis I $\times$ II plane is shown. In CASE-6, the axis I $\times$ III plane is also shown in addition. See Fig. 5 for the symbols.

区画は、第Ⅲ成分によってほぼ期待される順に序列化されている。この場合の第Ⅲ成分までの固有値の累積率は84.9%であった。

どのCASEにおいても、区画の配置はほぼ期待される順になっているが、CASE-5の中央部の区画は、ばらつきが他の区画より大きい。これは、二つの区画の全ての種の個体数が、それぞれ全区画の平均値に等しい場合、Rsnの計算のもととなる地点間相関指数Rsの値が、0/0(不定)になることに起因していると思われる。種の配置では、それぞれの種群がよくまとまった群となっており、CAでみられたような飛び離れた点が、どのCASEでもみられない。このことは、Rsn法では誤差の影響が抑えられていることを示すものである。またCASE-6, 7で、出現範囲が広い種群Eは、共通成分が大きい原点から遠い位置に配置されていることが、原点の近くに配置されるCAとは異なっている。

考 察

非確率的人工データの計算結果についてCAを用いた結果は、CASE-1の場合、もとのデータの1次元の構造を反映するという点では、Rsn法より優れている。これに対してRsn法では、1次元の構造の表示には相応しくないとされるアーチ効果^{1,5,6)}が大きい。これは、Rsn法では区画が単位球面上に配置されるためである。そこでそのことを考慮して、つぎのように考えることも可能である。3点の対等な関係は正三角形の頂点の位置で表わされるから、Fig. 4からは両端および中央部の、対等な三つの区画のグループが認識される。座標づけの目的が、環境勾配に沿って地点と種を並べることよりも、類似した地点と種の類別を行うことによってデータの構造的特徴を把握することであるとすれば、このような配置も意味がある。存在する環境勾配についての推察は、類別された種の実際の分布と類別された地点の特徴を注意深く吟味することによって行われるから、座標づけの座標軸が環境勾配の方向と一致することは、望ましいことではあるが絶対的とはいえないであろう。

Rsn法の座標づけがCAと異なる点の一つは、共通成分が示されることである。CAでは、共通成分が除かれているために、広範囲に出現する種の存在が区画と種の配置パターンに反映されない。Rsn法のもう一つの特徴は、CASE-3, 4で示されたように、類似した区画と種が大きくグループ化されることである。これは、環境勾配を推定するには誤解を与える恐れもあるが、この結果をクラスター解析するのには適している。さらにRsn法の利点として、CASE-1で述べたように、無意味な成分の変動が比較的小さいことがある。これに対してCAでは、無意味な成分の変動が、とくに第Ⅰ軸の

端部で大きい。このために有意な成分の取り方が、後に続くクラスター解析にも大きく影響する。

確率的人工データの計算結果についてCASE-7で示したように、CAもRsn法も、著しい優占種の影響を受けずに少数種の分布を反映するという点では、ほぼ同等である。しかしCAでは、その一方で少数種の不規則な変動を直接反映するため、有意な成分の固有値の累積率が低く、CASE-5, 6で示したように無意味な成分の変動が、とくに第Ⅰ軸の端部で大きい。このために実際のデータでは地点と種のグループが不鮮明になることがある。またCAでは、共通成分が除かれているので、広範囲に出現する種は原点の近くに配置され、このために不特定地点に散発的に出現する種との区別がされない場合がある。これは上の計算では取り上げなかったが、実際のデータで経験された。

したがってCAでは、このような少数種を除いて解析することが勧められている。¹³⁾しかし先にも述べたように、種の個体数の多さと種がもつ情報量の多さは一致しないから、一律に少数種を除くことには問題がある。また、不特定の地点に出現する種を予め識別することも難しい。CAでは、地点と種が同時に扱われるのでこれは大きな問題である。

さらにCAでは、顕著な種交代がなく総個体数が減少するような群集の変化は、検出し難いことがCASE-6の結果により示された。

一方Rsn法の座標づけでは、どのCASEでも少数種の不規則な変動の影響が抑えられており、クラスター解析に適した結果を与えている。これは、関係類似度Rsnおよびそのもとである地点間相関指数Rsの計算を通じて、不規則な変動の成分が取り除かれているためと考えられる。このために有意な成分の固有値の累積率が高く、無意味な成分の変動が比較的小さい。したがって有意な成分の取り方が後に続くクラスター解析に与える影響は、CAに比べると小さい。また広範囲に出現する種は、共通成分が大きい原点から遠い位置に配置されるので、不特定地点に散発的に出現する種とは区別される。

Rsn法では、地点と種がCAのように同時には扱われないので、情報を極力保存するために、地点の座標づけにはできるだけ全種を用い、不要な誤差を除くために、種の座標づけでは極端な少数種を除くようにすることができる。

さらにRsn法では、地点における総個体数の変化も、地点の座標づけに反映されることが、CASE-6の結果により示された。

CASE-5で示したように、Rsn法は全種が平均的に出現するような地点ではやや問題を生じるが、実際のデ

ータでは、そのようなことはあまりないと考えられる。

Rsn 法の利点とその位置づけ 群集のデータを解析し、各サンプルの特徴を客観的に示す方法として、多変量解析が優れていることは、すでに多くの研究者が認めている。⁸⁾ しかしながら、相関分析や主成分分析など、種間の関係や個体数の変動に一定の規則性を要求する数理解析法は、マクロベントスの個体数データのような場合、よい結果をもたらさない。そのような要求を必要としない数理解析法として、対応分析 (CA) やその変形である除歪法 (DCA) が近年評価されている。^{3,7)} しかしこの CA は、データの行および列和に対する比率を生起確率と考えて扱うことになるため、採集個体数が少ない種を多く含み、かつ不規則な変動が大きい群集の解析では、グループが正しく識別されないなどの問題を生じることがある。これに対して Rsn 法は、データ行列を直接解析するのではなく、関係類似度を介して間接的に解析するため、不規則な変動の影響を受け難く、類似した区画と種のグループをよく識別できる。このような間接的な方法は、主座標解析 (PCO) や多次元尺度解析 (MDS) と同様である。Rsn 法がそれらと異なる点は、これらの方法では、地点と種が別々に独立して扱われるのに対し、Rsn 法では、地点の座標づけに基づいて種の座標づけが行われることである。出現地点が限られた種を多く含むベントスのようなデータでは、地点間の関係の方が種間の関係よりも誤差の入り込む余地は少なく、また地点と種の対応もつけやすい。

Rsn 法の問題点として、地点グループ間の差異が誇張される可能性のあることが指摘されたが、それは、個別の種の分布を吟味しながら結果を解釈することによって防がれる。したがって Rsn 法は、群集変化のモニタリングや、海域の区分を目的とした調査などにおける一連の解析作業の 1 ステップとして、位置づけることができる。

謝 辞

本研究の機会と終始貴重なご助言を賜った、水産大学校網尾勝名誉教授に、また、論文作成の過程でご指導を頂いた、長崎大学水産学部玉置昭夫助教授および水産大学校増殖学科林健一教授の各位に対して、深甚なる謝意を表する。

文 献

- 1) 小林四郎：生物群集の多変量解析，第 1 版，蒼樹書房，東京，1995，pp. 71-147.
- 2) J. G. Field, K. R. Clarke, and R. M. Warwick: A practical strategy for analyzing multispecies distribution patterns. *Mar. Ecol. Prog. Ser.*, 8, 37-52 (1982).
- 3) 加藤和弘：生物群集の多変量解析とその地域環境計画への

応用。ランドスケープ研究，60, 46-55 (1996).

- 4) M. Kent and P. Coker: *Vegetation Description and Analysis*, 1st ed., CRC Press, California, 1992, pp. 293-303.
- 5) M. O. Hill and H. G. Gauch, Jr.: Detrended correspondence analysis, an improved ordination technique. *Vegatio*, 42, 47-58 (1980).
- 6) P. G. N. Digby and R. A. Kempton: *Multivariate Analysis of Ecological Communities*, 1st ed., Chapman & Hall, London, 1987, p. 97.
- 7) 加藤和弘：生物群集分析のための序列化手法の比較研究。環境科学会誌，8, 339-352 (1995).
- 8) B. J. Burd, A. Nemecek, and R. O. Brinkhurst: The development and application of analytical methods in benthic marine infaunal studies, in "Advances in Marine Ecology" (ed. by J. H. S. Blaxter and A. J. Southward), Vol. 26, Academic Press, London, 1990, pp. 221-234.
- 9) 石川公敏：沿岸域のマクロベントス群集の解析法。公害，24, 119-136 (1989).
- 10) 品川沙夫：底生動物相による海域環境解析の一方法。日本ベントス研究会誌，26, 49-65 (1984).
- 11) 品川沙夫：動物の種一個体数組成の類似度にもとづく多変量解析法とその適用例。下関女子短期大学紀要，10/11, 1-18 (1992).
- 12) 岩坪秀一：数量化法の基礎，第 1 版，朝倉書店，東京，1987，pp. 125-160.
- 13) H. G. Gauch, Jr.: *Multivariate Analysis in Community Ecology*, 1st ed., Cambridge Univ. Press, New York, 1982, pp. 144-160.

付 録

Rsn 法による地点と種の座標づけの概要

2 地点間の類似度 C'_x の定義 全調査地点での出現種数を S とすると、地点 X の種・個体数組成は、 S 次元の列ベクトル $\mathbf{x}$ で表される。同様に、地点 Y の種・個体数組成列ベクトルを $\mathbf{y}$ とすると、 $\mathbf{x}$ と $\mathbf{y}$ の類似度として、次の三つの因子を考えることができる。

- 1) 方向の類似度 $\gamma = \mathbf{x} \cdot \mathbf{y} / \|\mathbf{x}\| \|\mathbf{y}\|$
ただし、 $\cdot$ は内積を表すものとする。
- 2) 総個体数の類似度 $e_1 = 2\|\mathbf{x}\| \|\mathbf{y}\| / (\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2)$
- 3) 均等性の類似度 $e_2 = 2d_x d_y / (d_x^2 + d_y^2)$
ただし、 $d_x = \|\mathbf{x}\| / \|\mathbf{x}\| = 1$, $d_y = \|\mathbf{y}\| / \|\mathbf{y}\| = 1$,
 $\bar{\mathbf{x}}, \bar{\mathbf{y}}$ は、全成分がそれぞれの地点の 1 種あたり平均個体数に等しいベクトルとする。

これら三つの因子は、それぞれ独立に変化でき、かつそれぞれが群集組成の特性の側面を表しているから、 $\mathbf{x}$ と $\mathbf{y}$ の類似度を、これら三つの因子の積として定義することができる。 e_{2Y} は、木元の群集間類似度 C_x に一致するから、この三つの積を C'_x と定義する。

$$C'_x \equiv e_1 e_2 \gamma$$

個体数の基準化 種による個体数の相違が著しい場合、個体数の大きい種が重視される C'_x を、そのまま用いることは適切ではない。そこで、つぎのような基準化を行う。

第 i 種の, 出現した地点での個体数の平均を m_i とし, 地点 X と Y での個体数を, m_i で割った値を α_i, β_i とする。個体数の平均に全地点を用いないのは, 種ごとの個体数の大きさの基準を定めるのに, もともと生息しないかあるいは生息が確認されていない地点を含めるのは適切ではないと考えられるからである。この点が, データ行列の行和と列和の平方根で基準化する対応分析 (CA) との相違点の一つである。

地点間相関指数 R_s の定義 二つの, 独立な分布をする確率変数 x と y について, xy の期待値が, x と y それぞれの期待値の積に等しくなることを用いると, 種類数が十分多ければ, 上で述べた種ごとの平均個体数 m_i と, $\alpha_i \beta_i$ は独立と考えられるので,

$$y \doteq \sum m_i^2 \alpha_i \beta_i / \|x\| \|y\| \quad \Sigma \text{ は } i \text{ についての総和とする。}$$

そこで, α_i と β_i が独立な場合,

$$y_{ind} \doteq \sum m_i^2 \bar{\alpha} \bar{\beta} / \|x\| \|y\|$$

とできる。ただし, $\bar{\cdot}$ は全種の平均を表すものとする。したがって,

$$C'_n - C'_{ind} \doteq \varepsilon_1 \varepsilon_2 y - \varepsilon_1 \varepsilon_2 y_{ind} = \varepsilon_1 \varepsilon_2 \Delta$$

$$\text{ただし, } \Delta = \sum m_i^2 (\bar{\alpha} \bar{\beta} - \bar{\alpha} \bar{\beta}) / \|x\| \|y\|$$

となるから, 二つの地点の, 種・個体数組成が独立な場合を基準とした地点間相関指数 R_s を次式で定義する。

$\Delta > 0$ の場合:

$$R_s = (C'_n - C'_{ind}) / (1 - C'_{ind}) \\ \doteq \varepsilon_1 \varepsilon_2 \Delta / \{1 - \varepsilon_1 \varepsilon_2 (y - \Delta)\}$$

$\Delta < 0$ の場合:

$$R_s = (C'_n - C'_{ind}) / C'_{ind} \\ \doteq \Delta / (y - \Delta)$$

地点間関係類似度 R_{sn} の定義 地点 i と j の相関指数 R_{sij} を成分とする対称行列 $[R_{sn}]$ の, 第 i 列と第 j 列のベクトル R_i と R_j は, 地点 i および j と他の地点との関係を表している。そこで, 地点 i と j の, 他の地点との関係の類似度 R_{snij} を次式で定義する。

$$R_{snij} \doteq R_i \cdot R_j / \|R_i\| \|R_j\|$$

この操作により, R_{snij} を成分とする対称行列 $[R_{sn}]$ は正定値となり, 不規則な変動成分を減じ, 調査地点をグループ化することができる。

$[R_{sn}]$ の固有値分解による地点と出現種の座標配置できるだけ低いランク数 r で, $[R_{sn}]$ を最小二乗法の意味で最もよく近似する行列は, $[R_{sn}]$ の固有値分解により得られる。固有値の大きい順に, 大きさを固有値の平方根に等しくした r 個の固有ベクトルの第 i 成分を, 第 i 地点の座標とすれば, 地点を r 次元の座標系内に配置することができる。このとき, 二つの地点の位置ベクトル間の余弦が, 2 地点の R_{sn} の近似値を表し, 位置ベクトルの長さが, この座標系内で説明される割合を表している。一方, 地点座標系内で, 各地点における出現個体数を座標値とする種の位置ベクトルを, 長さが 1 となるように基準化したのちに, $[R_{sn}]$ の固有空間内に射影すれば, 全ての種を対等に, 固有空間内の点として座標配置できる。このとき, 種の位置ベクトルの各座標値は, 固有ベクトルの方向への, 種の寄与率を表す。