ファジイクラスター解析を用いた新しい生物群集解析法についての理論的考察

品川汐夫

Theoretical Considerations of New Strategy using Fuzzy Cluster Method for Analyzing Faunal Data

Sekio Shinagawa

Abstract

Shinagawa has proposed a new strategy using the ordination and fuzzy cluster methods for analyzing faunal data of marine macrobenthos and shown its effectiveness by numerical experiments on simulated data and analyzing actual faunal data. In the present paper, theoretical background of the proposed ordination method is investigated and compared with widespread methods, such as correspondence analysis (CA) and principal coordinate analysis (PCO). And also, significance of the application of fuzzy cluster analysis to fauual data based on the result of the ordination is indicated. The advantages of the new ordination method are that it is less affected by noises and conducts more information because the proposed indices, correlation and relational similarity between sites, work as a filter to remove noises inherent to faunal data. Contrary to these advantages, "arch effect" and meaningless common factor shown by the new method are pointed out as defects. Although these defects cause no problem in the following cluster analysis, caution should be taken in interpreting the result. Boundaries of the swarms of sites or species in the ordination space are usually unclear due to intrinsic continuity of species distribution along the environmetal gradients. The fuzzy clustering represents such community structures more precisely, and so makes it possible to detect transitional community. Moreover, summarization of varied distribution patterns of species using the fuzzy clustering is realistic and allows application of the other multivariate analysis to make a step for a factor analysis including environmental variables.

---

* 本論文は，1998年12月，長崎大学に提出した学位請求論文（学術博士）の一部を要約したものである．

1. 緒　言

海産マクロベントスは浮遊幼生期を除いて移動性に乏しく，底質環境の変化をよく反映することから，様々な環境モニタリングの中で重要な調査対象となっている．この調査の対象となる生物は定着型の多毛類から能動的に移動する十脚類まで，その移動能力は様々であり，大きさも，1～2 mmから数10mmの範囲にまで及ぶ．また分布のパッチ性が強く，ホトトギスガイのように1回の採泥で1000個体を超す濃密な群塊を形成する種から，潮流などにより偶発的に1個体だけ採集される種も多数存在し，主に種の生態的特性の違いに起因する個体数の相違が著しい．

このような不均質なデータの特性による数理解析の困難さは一般の生物群集解析でも指摘されることではあるが，海産ベントスの個体数データはとくにその傾向が強い．加えて，多くの場合船上から採泥されるため，定まった地点での定量採集が難しいことなど，サンプリング上の問題もあり，データの不規則な変動は陸上の植生のデータを上回るものであろう．

このようなデータの特性から，調査はできるだけ多数地点において継続的に行い，調査結果は，1地点での絶対値ではなく，「時空的広がりの中でのパターンを認識」することが重要とされる．[1]近年，経年的な継続調査が各地で行われるようになり，時空パターンを認識するのに必要な多くのデータが蓄積されつつある．しかし，蓄積されたこれらのデータが十分に活用されてはいない．[2]それは，大量の群集データから時空パターンを認識するための数理的方法が未だ十分確立されていないことが一因と考えられる．

植生学の分野においては，1950年代の初めから，数理解析の一方法として群集の座標づけが行われてきた．これは，群集を環境傾度に沿って座標空間に布置することにより，視覚的に多くの情報を得て仮説を導こうとするものである．[3]1970年代に入ると，コンピュータの普及により数値分類法としてのクラスター解析が広く応用されるようになり，同時にそれまでの主観的な方法に替わる数学的な最適化に基づく座標づけ法が急速に発展し，両者は植生学の分野での群集解析に大きな貢献をしてきた．[4,5]座標づけの主な方法として，主成分分析法（PCA），主座標解析法(PCO)，非計量的多次元尺度解析法（NMDS），対応分析法（CA）などがある．中でも対応分析法（CA）は，その後の多くの座標づけ法の発展の基礎となった，重要な方法とされる．[3]植生学の分野では，Roux and Roux[6]が最初にこれを用いたとされ，その後多くの適用例が続く中でアーチ効果や収縮効果などの欠点[4]が指摘されるようになり，このような欠点を取り除くためにHill[7]とHill and Gauch[8]は除歪法（DCA）を開発した．さらにHill[9]は，CAによる座標づけを用いて指標種を定め，その在不在によって地点を順次2分割していく分割型のクラスター解析法（TWINSPAN）を開発し，植生学の群集解析では普及している．[3,10]我が国においても，Kobayashi[11]が群平均法の欠点を取り除く独自のクラスター解析法を提案

し，また加藤[12,13]が座標づけの諸方法の比較研究をするなど，この分野での研究が進められている．

一方，これらの方法の海産ベントス群集への適用は，類似度によるデンドログラムの作成のほかは適用例が少なく，数理解析法についての研究は，海外ではいくつかの例が見られるが，[14−17]我が国では赤嶺[18]と石川[19]があるほか、底生魚類群集の解析に Williams and Stephenson[20]の方法を適用した松宮，和田[21]がみられる程度である．その理由の一つに，上に述べた海産マクロベントス群集データの特性のために，植生学の分野で発展した諸方法のこの分野への直接的適用が，必ずしもよい結果をもたらさないことが挙げられる．植生学の分野で適用例が多いDCA について Field *et al.* [22]は，それがアーチ効果を強制的に取り去るために意味のある変動まで消し去ることもあることから，海産のベントス群集の解析では奨められないとしている．したがって，この分野においては，独自の解析方法の研究が必要である．

一方，品川[23]は，不均質かつ不規則な変動が大きい海産マクロベントス群集の個体数データに適した座標づけと，ファジイクラスター解析を用いた新しい群集の数理解析法を提案した．品川，多部田[24]は，マクロベントス群集に模して作成した人工データによる数値実験を行い，この方法がノイズの影響を受け難く，したがってマクロベントスのような群集のデータ解析に適していることを示した．また，品川，多部田[25]は，過去に公表されているマクロベントスの調査データを用いてこの方法と対応分析法（CA）の比較を行い，新座標づけ法に基づくクラスタリングは誤分類が少なく，群集構造がより正確に把握できること，およびファジイクラスター解析が，群集の連続的変化を表す上で極めて効果的であり，通常のクラスター解析では捉え難い，推移帯の群集（ecotone）も正しく位置づけられることを明らかにした．さらに，品川，多部田[26]は，熊本港周辺の干潟で1991年を除く1982年から1996年の過去14年間，継続して行われたマクロベントス調査データにこの方法を適用し，この方法が，人為的環境改変に伴う群集構造の変化とその要因を解明する上でも効果的な，優れた方法であることを示した．

本論文では，この方法の基礎となる新座標づけ法の理論的背景と他の諸方法との理論的比較，およびこの方法を適用するさいの理論的問題点などについて考察した．また，群集解析にファジイクラスター解析を適用する意義などについて検討を加えた．なお，具体的方法については，すでに本誌[27]および他誌[23]において公表しているが，参考のために末尾の付録Ⅰ～Ⅲにその概略を紹介した．

2. 品川[23]の提案した新座標づけ法についての理論的考察

2・1 行列の固有値分解による座標づけ法について

多変量データを空間に布置する方法と行列の固有値分解または特異値分解との関係について

は多くの文献がある．[28~30]ここではその中から，新座標づけ法の理論的根拠となるものを整理して，主成分分析法（PCA），主座標解析法（PCO），対応分析法（CA）などとの理論的相違や類似点を明らかする．また，新座標づけ法に特有の理論的問題点について考察を行う．

2・1・1　対称行列の最良近似

s_{ij} を第 i，j 成分とする $n\times n$ の半正定値対称行列を S，a_i を第 i 成分とする n 次元の列ベクトルを a とし，行列演算 aa' により S を最良に近似する行列を，$I=\sum_{i=1}^{n}\sum_{j=1}^{n}(s_{ij}-a_ia_j)^2$ を最小にするという基準により求めることを考える．I を行列演算で表すと，

$$
\begin{aligned}
I &= \mathrm{trace}\ \{(\mathbf{S}-\mathbf{aa'})'(\mathbf{S}-\mathbf{aa'})\}\\
&= \mathrm{trace}\ \{\mathbf{S}^2-\mathbf{Saa'}-\mathbf{aa'S}+(\mathbf{a'a})\,\mathbf{aa'}\}\\
&= \mathrm{trace}\ (\mathbf{S}^2)-2\mathrm{trace}\ (\mathbf{Saa'})+(\mathbf{a'a})\ \mathrm{trace}\ (\mathbf{aa'})\\
&= \mathrm{trace}\ (\mathbf{S}^2)-2\mathbf{a'Sa}+(\mathbf{a'a})^2
\end{aligned}
\qquad \ldots(1)
$$

ただし，trace は行列の対角項の総和，$'$ は行と列の転置を意味する．

したがって，

$\dfrac{\partial I}{\partial \mathbf{a}}=-4\mathbf{Sa}+4(\mathbf{a'a})\,\mathbf{a}=0$ とおくことにより，$\mathbf{Sa}=(\mathbf{a'a})\,\mathbf{a}$ を得る．

このことから，aa' により S を最良に近似するベクトル a は，その大きさが固有値の平方根に等しい S の固有ベクトルであることが分かる．したがって，新座標づけ法は，関係類似度によって表された地点間の関係を最も良く表す地点の布置を求める方法ということができる．

2・1・2　行列 Y'Y の固有値分解と座標づけ法

p 次元空間における n 個の点の座標を成分とする $p\times n$ の座標行列を Y とする．

$\mathbf{Y}=(\mathbf{y}_1,\mathbf{y}_2,\cdots,\mathbf{y}_n)$，　ただし，$\mathbf{y}_i$ は点 i の座標列ベクトル．　...(2)

このとき，点 i と j のユークリッド距離を d_{ij} とすれば，

$d_{ij}^2=(\mathbf{y}_i-\mathbf{y}_j)'(\mathbf{y}_i-\mathbf{y}_j)=\|\mathbf{y}_i\|^2+\|\mathbf{y}_j\|^2-2\mathbf{y}_i'\mathbf{y}_j$　...(3)

したがって，$n\times n$ 行列 Y'Y の第 i，j 成分を s_{ij} とすると，

$d_{ij}^2=s_{ii}+s_{jj}-2s_{ij}$　...(4)

となる．

対称行列 Y'Y は半正定値であるから，そのランク数を r（$r\leq \min(p,n)$）とすれば，大きさが固有値の平方根に等しい r 個の n 次元固有ベクトルを行とする $r\times n$ の座標行列 F によって，Y'Y=F'F と表される．このことは，F で表される r 次元の空間内における n 個の点間距離は，それぞれもとの Y で表される p 次元空間内の点間距離に等しいことを示す．すなわち，行列 Y'Y の固有値分解は，Y で表される p 次元空間内の n 個の点群を，その各2点間の距離を変え

ることなく，Yのランク数に等しい低次元の空間内へ移す方法といえる．このとき，n個の点のr次元空間における第i座標の二乗和は第i固有値に等しくなっているから，固有値が大きい順に選んだk（$k<r$）次元の空間は，同じ次元数の空間で，原点からの距離の二乗和が最大となる空間である．

関係類似度行列Rsnは，もととなる相関指数行列Rsとその第i列（または行）成分の二乗和を第i対角成分とする対角行列Nによって次のように表される．

$$\mathrm{Rsn} = \mathrm{N}^{-1/2}\mathrm{Rs}^2\mathrm{N}^{-1/2} \qquad \ldots (5)$$

$\mathrm{Y=RsN}^{-1/2}$とおけば$\mathrm{Rsn=Y'Y}$となるから，新座標づけ法は座標行列$\mathrm{RsN}^{-1/2}$の各列で表される点を，2点間の距離を変えずにr次元の球面上に布置し，さらに各点が原点からできるだけ離れた位置に布置される低次元の平面を求める方法ということができる．

2・1・3 主成分分析法（PCA）および対応分析法（CA）との関係

2・1・2においてYを生のデータ行列Xとすれば，中心化（平均値に原点を移動）とスケール変換を伴わない主成分分析法となる．また，中心化を行えば分散行列に基づく主成分分析法となり，さらに分散が1となるようスケール変換を行えば相関行列に基づく主成分分析法となる．

したがって，新座標づけ法は，相関指数行列Rsをデータ行列とみなし，これを$\mathrm{N}^{-1/2}$でスケール変換した，中心化を行わない主成分分析といえる．一方，

$$\mathrm{Y} = \mathrm{D}_r^{-1/2}(\mathrm{X} - \mathrm{X}_{ind})\,\mathrm{D}_c^{-1/2} \qquad \ldots (6)$$

とおけば，これは対応分析法（CA）となる．[3]ただし，D_rとD_cは，それぞれデータ行列Xの行和と列和を対角成分とする対角行列である．

また，X_{ind}は，Xの行と列が互いに独立として，その行和と列和から推定されるXの期待値で，次式で与えられる．

$$\mathrm{X}_{ind} = \mathrm{D}_r \mathrm{1}_p \mathrm{1}_n{}' \mathrm{D}_c / x_{..} \qquad \ldots (7)$$

ただし，$x_{i.} = \sum_{k=1}^{n} x_{ik}$ ，$x_{.j} = \sum_{k=1}^{p} x_{kj}$ ，$x_{..} = \sum_{i=1}^{p}\sum_{j=1}^{n} x_{ij}$

$\mathrm{D}_r = \mathrm{Diag}\,[x_{i.}]$ ，$\mathrm{D}_c = \mathrm{Diag}\,[x_{.j}]$

$\mathrm{1}_n$，$\mathrm{1}_p$は，すべての成分が1に等しい，それぞれn次，p次の列ベクトルであり，$\mathrm{Diag}\,[x_{i.}]$，$\mathrm{Diag}\,[x_{.j}]$は，それぞれ$x_{i.}$，$x_{.j}$を第i，j対角成分とする対角行列である．

この場合のY'Yの固有値分解によるn個の地点の布置では，全地点の原点からの距離の二乗和に$x_{..}$を乗じた値は，もとのデータ行列Xの行と列の独立性を検定する統計量χ^2に等しくなる．対応分析法（CA）では，この座標値をさらに対応する列和の平方根で割って，2点iとj間の距離の二乗が次式となるように変換される．

$$d_{ij}^2 = \sum_{k=1}^{p} \frac{1}{x_{k.}} \left(\frac{x_{ki}}{x_{.i}} - \frac{x_{kj}}{x_{.j}} \right)^2 \qquad \ldots (8)$$

したがって，対応分析法（CA）による地点の布置は，種の総個体数の逆数で重みづけした，地点における各種の組成比率の総合的相違を表現している．**Y** を **Y'** に置き替えれば，これはそのまま種の座標づけとなる．

対応分析法における **Y'Y** を地点間の類似性を表す行列とみなせば，類似性行列の固有値分解という点で新座標づけ法との共通点がみられるが，次のような相違点も浮かび上がる．

(1) 類似度を求めるデータの基準化

対応分析法（CA）では，もとのデータを行和（対応する種の総個体数）と列和（対応する地点の総個体数）の平方根で基準化する．一方，新座標づけ法における地点間相関指数 Rs の計算では，行については，その種が出現した地点における個体数の平均で基準化する．また，列については列ベクトルのノルムで基準化する．その結果，行の基準化において対応分析法は総個体数に比べて出現地点数が少ない種の重みを重くし，出現地点数が多い種の重みを軽くする．例えば 4 地点での出現個体数が（1,1,1,1）の種と（1,1,0,0）の種を行について基準化すると，前者では出現地点に $\frac{1}{2}$ の重みが，後者では $\frac{1}{\sqrt{2}}$ の重みが与えられる．一方，新座標づけ法ではどちらにも **1** の重みが与えられる．

分布が広い種よりも限られた種の方が地点の特性をよく表すと考えれば対応分析法の方法に理があるが，マクロベントスのデータでは，限られた地点で特定の種が偶然大量に採集される場合があり，その場合，このような偶然変動が座標づけに反映され易くなる．

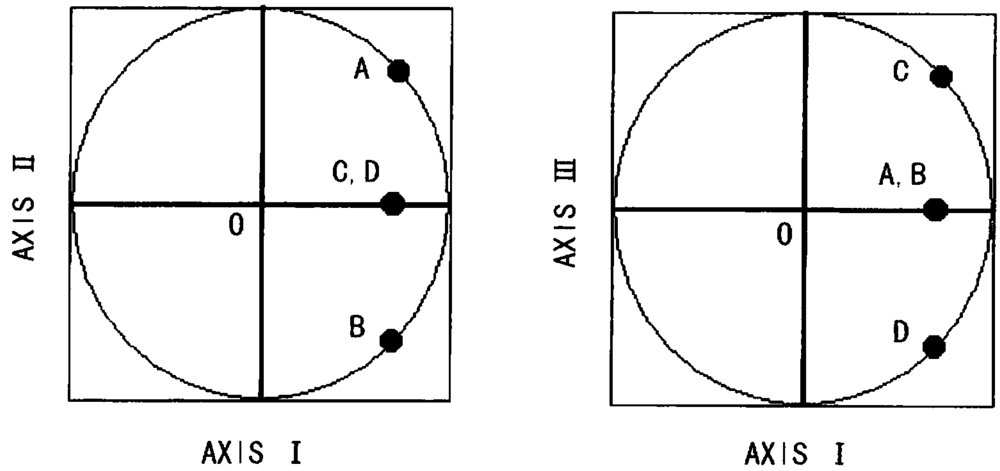
また，列については，対応分析法（CA）は行と同じ基準化を行うため，行（種）と列（地点）に双対性がある．一方，新座標づけ法にはそのような双対性がなく，類似度行列の対角項を1に揃えるため，地点の布置が球面上に拘束される．その結果不規則な変動は抑えられるが，平面内では円周状の布置となり，アーチ効果[4]を生じる．

(2) データの中心化

対応分析法（CA）による地点の布置では，**X** の行（種）と列（地点）が独立と仮定した場合に期待される種組成比率すなわち行和の分布が基準となり，そのような種組成の地点の位置が原点となる．それはまた，総個体数で重みづけした地点の平均座標に一致する．したがって，地点の位置の原点からの距離は，平均的な種組成比率からの偏りを表す．この結果，点の布置が示す生態学的な意味は明瞭となるが，偶然大量に採集された種が存在する場合，原点の位置がそのような種の影響を強く受ける．また，地点当たり総個体数の相違は示されないなどの問題点もある．

一方，新座標づけ法においては，各 2 地点間で種の出現個体数が独立な場合を基準として地点間相関指数を計算するので，全地点に共通した基準はなく，原点の中心化は行われない．この場合，各地点の位置の原点からの距離は，その平面で地点間の関係が説明される度合いを示す．また，地点当たり総個体数の相違は Rs の計算に反映されている．

原点を中心化しない場合



原点を中心化した場合

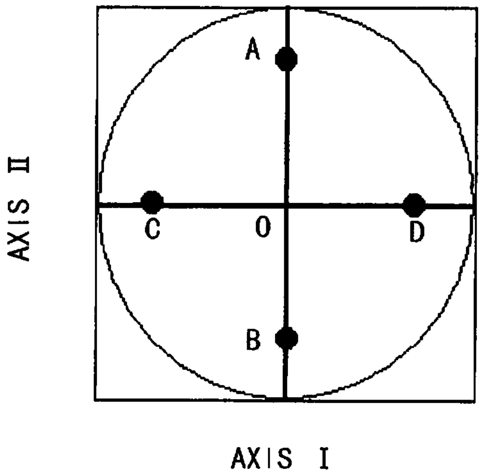


図１．原点を中心化しない場合（上）とした場合（下）の座標づけの比較．

原点を中心化しない場合（上），共通成分（AXIS　Ⅰ）が表示されるので，広域分布種の存在が座標づけに反映されるが，地点間の相違を示すためには，中心化を行った場合（下）より多くの平面（座標づけ空間の次元）が必要である．

中心化を行った場合，地点は平均からの位置の変動が最大となる平面に布置されるので，不規則な変動に対しても敏感になる．また，共通成分が除去されるので，全域に広く分布する種の存在は地点の布置パターンに反映されない．新座標づけ法による地点の布置は，そのパターンに広域分布種の存在も反映するが，地点間の関係を表すのに必要な空間の次元数が，中心化を行った場合より大きくなる（図１）．

2・1・4 主座標解析法（PCO）との比較

主座標解析法（PCO）は，地点間の相違度が与えられた場合，点間距離が相違度に等しくなるように地点を布置する方法で，計量的多次元尺度解析ともいわれる．[30]点の座標は，相違度の二乗を成分とする行列を D とすると，$-\frac{1}{2}$D の左と右から直交射影行列 Q を乗じて D_n とし，D_n を固有値分解することにより求められる．ただし，Q は次式で与えられる行列で，行列を中心化して，平均からの偏差行列に変換する働きをする．

$$Q = (I - 1_n1_n' / n) ,\quad I \text{ は } n \times n \text{ の単位行列} \qquad \ldots(9)$$

$$D_n = -\frac{1}{2} QDQ \qquad \ldots(10)$$

地点間の関係として類似度が与えられた場合，$\frac{1}{2}$D の成分に類似度の1の補数を用いればよい．ところが，このようにして得られた行列を中心化した結果は，もとの類似度行列 S を中心化した結果と完全に一致する．なぜなら，

$Q1_n = 0$だから，$-Q(1_n1_n' - S)Q = QSQ$ となるからである．

このことから，Rsn の固有値分解による座標づけと主座標解析法（PCO）の相違は，中心化を行うか否かによることが分かる．中心化を行った場合と行わない場合の座標づけの相違についてはすでに上に述べた．

2・2 関係類似度の問題点についての考察

上では，関係類似度行列を固有値分解する利点として共通成分が表示されることを挙げた．しかし，一次元的な群集構造が長く連続する場合，たとえ広域分布種が存在しなくとも，関係類似度行列の共通成分が大きくなる場合が有り得る．このことについて理論的な考察を行う．

$n \times n$ の地点間の類似度を表す行列 R が，対角部分は全成分が1である $m \times m$ の小行列 g 個からなり，他の成分がすべて等しく β となっている場合を考える．これは，同じ群集内では種・個体数組成が完全に一致し，異なった群集間では類似度が β に等しい g 個の群集が一次元的に連続していることを意味している．小行列を A とすると，

$$R = \begin{pmatrix} A & & \beta \\ & A & \\ & & \ddots & \\ \beta & & & A \end{pmatrix}, \qquad A = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} = 1_m1_m' \qquad \ldots(11)$$

R を変形すると，

$$\mathbf{R} = (1-\beta)\begin{pmatrix} \mathbf{A} & & \mathbf{0} \\ & \mathbf{A} & \\ \mathbf{0} & & \ddots \ \mathbf{A} \end{pmatrix} + \beta \mathbf{1}_n \mathbf{1}_n' \qquad \ldots(12)$$

したがって，

$$\mathbf{R1}_n = m(1-\beta)\mathbf{1}_n + n\beta\mathbf{1}_n = \{m(1-\beta) + n\beta\}\mathbf{1}_n \qquad \ldots(13)$$

となるから，$\mathbf{1}_n$ は行列 R の固有値 $\lambda_0 = m(1-\beta) + n\beta$ に対応する固有ベクトルで，共通成分を表している．さらに，$\sum_{i=1}^{g} a_i = 0$ となるように a_i を選び，

$$\mathbf{a} = [a_1\mathbf{1}_m', a_2\mathbf{1}_m', \cdots, a_g\mathbf{1}_m']' \qquad \ldots(14)$$

とすると，

$$\mathbf{A1}_m = m\mathbf{1}_m \text{ および } \mathbf{1}_n'\mathbf{a} = 0 \text{だから，} \quad \mathbf{Ra} = m(1-\beta)\mathbf{a} \qquad \ldots(15)$$

したがって a は，固有値 $\lambda_1 = m(1-\beta)$ に対応する R の固有ベクトルである．互いに直交するこのようなベクトル a は一意には定まらないが，$g-1$ 通り選ぶことができ，それらは全て同じ固有値 $m(1-\beta)$ をもつ．R のランク数は g だから，これで行列 R の全ての固有値と固有ベクトルが求まったことになる．

以上のことから，それぞれ m 個の地点を含む異なった群集が g 個，一次元的に連続する場合，行列 R のランク数は g で，一つの共通成分と，$g-1$ 個の群集を識別する成分をもつ．また，群集を識別する成分は共通成分に垂直な平面内で，原点の中心化が行われている．

$\beta>0$ の場合，共通成分は常に識別成分より大きくなる．したがって，一般の類似度（$\beta>0$）では，地点間の関係を表すには共通成分を除いた方がよい．一方，$\beta<0$ の場合，共通成分は識別成分より小さくなるが，

$$\beta < \frac{m}{n-m} = -\frac{1}{g-1} \qquad \ldots(16)$$

の場合には，共通成分の固有値 $\lambda_0<0$ となるため，R で示される地点間の関係を表現する座標布置が不可能になる．相関指数 Rs から関係類似度 Rsn を計算する理由の一つはこのためである．

異なった群集に属する地点間の相関指数が全て α の場合，関係類似度行列は上の R で，

$$\beta = \frac{\alpha^2(n-2m)+2m\alpha}{m+(n-m)\alpha^2} = \frac{\alpha^2(g-2)+2\alpha}{1+(g-1)\alpha^2} \qquad \ldots(17)$$

と置いたものとなる．したがって，共通成分および識別成分の固有値の全成分 n に対する割合（寄与率）はそれぞれ，

$$\frac{\lambda_0}{n} = 1 - \frac{(g-1)(1-\alpha)^2}{g\{1+(g-1)\alpha^2\}}, \qquad \frac{\lambda_1}{n} = \frac{(1-\alpha)^2}{g\{1+(g-1)\alpha^2\}} \qquad \ldots(18)$$

となる．$g=2\sim5$ の場合の$\dfrac{\lambda_0}{n}$および$\dfrac{\lambda_1}{n}$と α との関係を図２に示す．図では固有値の割

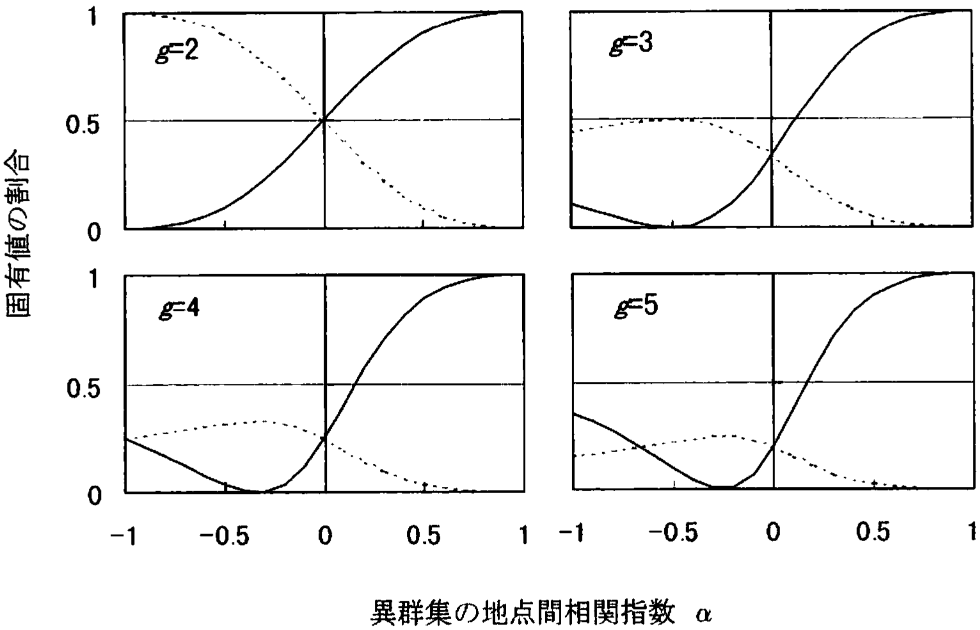


図２．異群集の地点間相関指数と関係類似度行列の固有値寄与率との関係．

g 個の異なった群集が１次元的に連続し，同じ群集内では種・個体数組成が完全に一致，異群集の地点間では相関指数 $Rs=\alpha$ の場合の関係類似度行列各固有値の，全成分に対する割合（寄与率）を示す．
実線；共通成分の固有値，点線；識別成分の固有値，g；異なった群集の数．

合を，共通成分は実線で，また識別成分は点線で示している．図から，$\alpha>0$の場合にはいずれの場合も共通成分が最大の成分になることが分かる．そのような場合には広域に分布する種が存在すると解釈してもよい．しかし，$\alpha<0$の場合，共通成分は $\alpha\fallingdotseq-\dfrac{1}{g-1}$ において極小となり，それより α が小さくなると共通成分はかえって増大する．すなわち，群集間の差異が大きくなるほど共通成分が大きくなるという矛盾を生じる．とくに $g\geqq5$では，$-1<\alpha<0$の範囲内で共通成分の方が識別成分より大きくなる範囲を生じる．実際のデータでは誤差成分が加わるため，$g\leqq4$であってもこのような逆転が起こりうる．

共通成分の存在はクラスタリングには影響しないから，クラスタリングを目的とした座標づけにおいてはこのことが問題とはならないが，座標布置から群集構造を推察する際には問題となる．したがって，新座標づけ法で座標づけした場合，共通成分の解釈は実際のデータに立ち返って慎重に行う必要がある．

3. ファジイクラスター解析の群集解析への応用について

3・1 群集の連続性とファジイクラスタリングの意義

群集解析における座標づけとクラスター解析は，それぞれ群集連続体説と群集個別体説に対応づけられる．[10,31]今日では，多くの研究者が連続体説に傾いており，群集に明瞭な境界を定めることはできないとされる．[31]しかし，それでもなおクラスター解析は有効である．群集を座標づけした結果から連続的な環境傾度が推定できることはむしろ困難な場合が多く，多かれ少なかれいくつかのグループが認識されるのが普通である．[3]したがって，クラスター解析によってグループの特徴を明らかにし，それに基づいて背後にある生物的，非生物的要因を推察することが効果的である．小林[10]は，座標づけされた点がどの程度密であれば不連続な単位とみなし，どの程度散らばっていれば連続体とみなすかという画然とした区別はなく，同じデータを分類することも座標づけすることも可能であるとしている．

一方，ファジイクラスター解析が扱うファジイ集合は，統計的なランダム性に起因しないあいまいさを扱うために Zadeh[32]によって提案されたもので，確率が事象の生起に関するあいまいさを表しているのに対し，ファジイ集合は事象の意味のあいまいさを表しているとされる．[33]このような集合の概念は，まさに上に述べた群集を扱う概念として相応しい．連続性を認めつつ，自然なクラスターを見分ける方法として，群集解析にファジイクラスター解析を適用することの重要な意義がある．

しかしながら，現実には，群集の解析にファジイクラスター解析を適用した例は，筆者の知る限り Bezdek[34]のほかにはない．Bezdek[34]は環形動物群集の解析に適用した例を紹介しているが，これは88種を20地点の座標軸に直接布置したものである．先に述べたベントスのデータの特性から，このように直接座標づけをすると多くの場合，意味のあるクラスターが形成されない．Bezdek[34]の例においても，特異な分布をする種がクラスターとして識別されているだけである．ファジイクラスタリングが本質的に適しているにも関わらず普及していない理由の一つとしてこのようなことが考えられる．

ここで提案した，座標づけをファジイクラスター解析の前処理として用いることは，この点の解決策である．その根拠は，座標づけが数学的基準の最適化を行う結果，不規則な変動成分を除去するフィルターとしての役割を果たすことにある．

ここで用いたファジイ k-means 法は，クラスターの形状が塊状であることが前提である[35]が，新座標づけ法による座標づけは，座標空間内に塊状のクラスターを形成する傾向がある[24]から，ファジイ k-means 法の前提にも合致している．

3・2 クラスタリング結果を評価する方法について

ファジイクラスタリングの利点に，所属率を用いて結果の評価ができることがある．Roubens[36]はそのための指数として，分割係数（partition coefficient），分割度（non fuzziness index），あいまい度(fuzziness performance index)，分類エントロピー(normalized classification entropy)などを提案している．これらはいずれも分類しようとする対象の空間布置の状況からクラスタリング結果を評価するものである．新座標づけ法では，地点と種の座標空間への布置を，直接もとのデータ行列からではなく，関係類似度行列を介して行う．したがって，クラスタリング結果の評価は，座標布置の状況よりも，もとのデータに基づいて行う方がよい．そこで，品川，多部田[25]は，もとのデータに示された種の分布と地点のクラスタリングを関連づけ，クラスタリング結果を評価する方法として，つぎのような指数を提案した（巻末付録Ⅲ参照）．

1) 種の平均地点群所属率と分布のあいまい度および地点のクラスタリングのあいまい度
2) 種群と地点群の平均的関係および種群の分布のあいまい度
3) 種群の個体数および種類数とその特化係数[37]

個別の種の個体数は不規則な変動が大きいため，その分布から環境傾度を明らかにすることは一般に困難である．しかし，分布の類似した種について平均化すれば，これらの不規則成分が相殺されて，環境傾度に対応した変動パターンを見出せることが期待される．上に提案した指数はそのような平均化を行うものであり，多様な種の分布パターンから情報を集約する現実的な方法といえる．通常のクラスター解析ではこのような集約は難しく，加藤[13]は，地点と種の群の対応がつけ難いことをクラスター解析の欠点として挙げている．ファジイクラスター解析は，このような欠点を補うものである．

3・3 クラスター数と重み指数を定める問題

一般にファジイクラスタリングでは，クラスター数 g をあらかじめ定めなければならない．さらに，ファジイ k-means 法では，重み指数 m も定めておく必要がある．McBratney and Moore[38]は，g と m を同時に定める方法として，クラスター内分散 J_m の m に関する変化率の絶対値が最大となる値に g を乗じた値が最小となるように定める方法を提案している．しかし，そのためには多数の試行計算が必要となる．実際には経験的に m=2の付近で J_m の変化が最大となることが知られており，m=2とすることによる問題点は指摘されていない．

クラスター数 g の定め方については，3・2に挙げたいくつかの指数を用いることが考えられる．[38]しかし，実際問題では，分割係数，分割度が小さく，あいまい度が大きくなっても，クラスター数を増加した方が妥当な解釈ができる場合あるいはその逆も生じる．したがって，

クラスター数は，解析者の経験と対象としている問題に関する知識に基づいて，何回かの試行計算結果の観察から定めるしか方法はない．指数はその際の参考とされる．Bezdek[34]は，クラスター数を変えたときの所属率が変化する様子から，クラスターの意味について多くの情報を得ることができることを示している．

4．結　言

4・1　新座標づけ法の利点と問題点

理論的考察により明らかにされた新座標づけ法の利点はつぎの点にある．

1）新座標づけ法は，関係類似度行列で表された地点間の関係を低次元の空間で最もよく表す座標を与える．それが対応分析法（CA）などと異なる点は，原点の中心化を行わないことである．このことにより広域に分布する種の存在が，地点の布置パターンに反映されるので，中心化を行った場合より多くの情報が表される．

2）新座標づけ法では，地点間相関指数 Rs および関係類似度 Rsn の計算を通じて不規則な変動の影響が取り除かれ，また Rs の計算では，すべての種に同じ重みが与えられる．したがって，マクロベントスのような不均質で不規則な変動が大きいデータから有意な情報を抽出するのに適している．これに対して対応分析法（CA）では，データ行列を直接解析し，また出現地点の少ない種に大きな重みが与えられるため，偶然大量に出現する種が座標づけに直接影響し，不規則な変動の影響が大きくなる．

3）マクロベントス調査では，地点当たり総個体数の相違も重要な意味をもつと考えられるが，対応分析法（CA）による座標づけにはそれが反映されない．一方，新座標づけ法においては，それは地点間相関指数 Rs の計算に反映されるので，汚染源の近くにおける総個体数の減少等の変化も検出できる．

一方，新座標づけ法を適用する上での問題点はつぎのとおりである．

1）地点が単位球面上に布置されるので，一次元的な構造の表現には相応しくないとされるアーチ効果が大きい．

2）実際の群集構造には関係のない無意味な共通成分を生じることがある．

これらの問題点は，クラスター解析の結果には影響しない．実際問題では，座標づけにより示された軸がどのような環境傾度に対応しているかということよりも，それぞれの地点群でどのように群集の組成が変化しているかということの方が本質的である．したがって，誤分類を生じ難い座標づけであることの方が重要である．上の問題点は，つぎのファジイクラスター解析により，種群および種群を構成する種の分布特性と，地点群の関係を明らかにすることによって解決される．

4・2　群集解析にファジイクラスター解析を適用する利点

群集解析にファジイクラスター解析を適用する利点はつぎの点にある．

1）環境傾度に沿った群集の変化は連続的である．このために通常のデンドログラムによるクラスター解析では推移帯の群集（ecotone）を識別し難い．また識別できたとしても，他の地点群との関係を位置づけることが困難である．新座標づけ法による座標づけとそのファジイクラスタリング結果を用いれば，このような群集も正確に把握できる．

2）一般に地点の座標づけは種の座標づけに比べて群が比較的明瞭であるが，種の座標づけは地点の場合に比べると群の境界が不明瞭で，群の広がりが大きくなる．これには採集地点の設定や採集誤差も影響していると考えられるが，種による分布パターンの多様さを反映しているとみることもできよう．このような種の分布を通常のクラスター解析によって要約することは適当でない．しかし，個々の種の分布を取り上げても複雑さが増すばかりである．ファジイクラスター解析による種群の識別は，より現実的であり，ファジイクラスター解析の結果から導かれるさまざまな指数は，地点群と種の分布の対応関係を簡潔に表現し，群集の特徴とその連続的変化を正確に把握するのに役立つであろう．

3）個別の種の個体数は不規則変動が大きいため，環境要因と直接関係づけることは一般に困難である．しかし，分布特性の類似した種の個体数の和を用いれば不規則変動が緩和され，平均的な意味で種の分布と環境要因とを関係づけることができると期待される．ファジイクラスター解析によって得られた種群の個体数は，そのような解析を可能にし，環境の改変に伴う群集の変化を予測する解析への道を拓くことができる．

5．謝　辞

本研究の端緒を与えて頂き，さらに論文の作成に際しては懇切丁寧なご指導とご助言を賜り，論文を御校閲いただいた，長崎大学教授　多部田修博士，および御高閲いただいた長崎大学教授　崎山毅博士，同教授　夏刈豊博士，同教授　松野健博士，同助教授　玉置昭夫博士，同助教授　山内淳博士に深甚なる感謝の意を表する．

また，水産大学校名誉教授　網尾勝博士には，ベントスの採集方法から種の査定，分析結果の解釈に至るまで細かくご指導を賜り，本研究の動機と方向性を与えていただいた．ここに深い感謝の意を表する．

本研究の基礎は，（株）下関理化学分析センターにおけるベントスの分析業務の中で形成された．ここに記して，同センター社長　和田照人氏並びに職員各位に対して深く謝意を表する．

文 献

1) 菊池泰治：海域における富栄養化と底棲生物の指標性，沿岸海域の富栄養化と生物指標（日本水産学会編），恒星社厚生閣，東京，1982，pp. 84～100.

2) 小倉紀雄，高田秀雄，向井 宏：沿岸の汚染機構，海洋環境を考える（日本海洋学会編），恒星社厚生閣，東京，1994，pp. 3～12.

3) Kent, M. and P. Coker : Vegetation Description and Analysis, 1st ed., Belhaven Press, London, 1992, pp. 162～319.

4) Gauch, H.G. : Multivariate Analysis in Community Ecology, 1st ed., Cambridge Univ. Press, New York, 1982, pp. 109～172.

5) Digby, P.G.N. and R.A. Kempton : Multivariate Analysis of Ecological Communities, 1st ed., Chapman & Hall, London, 1987, pp. 49～111.

6) Roux, G. and M. Roux : A propos de quelques methodes de classification en phytosociologie, *Revue de Statistique Appliquee*, **15**, 59～72 (1967).

7) Hill. M.O. : DECORANA-a Fortran program for detrended correspondence analysis and reciprocal averaging, Cornell Univ., Ithaca, 1979.

8) Hill, M.O. and H.G. Gauch : Detrended correspondence analysis: an improved ordination technique, *Vegitatio*, **42**, 47～58 (1980).

9) Hill, M.O. : TWINSPAN, a Fortran program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes, Cornell Univ., Ithaca, 1979.

10) 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.

11) Kobayashi, S. : Heterogeneity ratio: a mrasure of beta-diversity and its use in community classification, *Ecol. Res.*, **2**, 101～111 (1987).

12) 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）.

13) 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）.

14) Field, J.G., K.R. Clarke, and R.M. Warwick : A practical strategy for analyzing multispecies distribution patterns, *Mar. Ecol. Prog. Ser.*, **8**, 37～52 (1982).

15) Gray, J.S., M. Aschan, M.R. Carr, K.R. Clarke, R.H. Green, T.H. Pearson, R. Rosenberg, and R.M. Warwick : Analysis of comminity attributes of the benthic macrofauna of Frierfjord/Langesundfjord and in a mesocosm experiment, *Mar. Biol. Prog. Ser.*, **46**, 151～165 (1988).

16) Ardisson, P.L., E. Bourget, and P. Legendre : Multivariate approach to study species assemblage spatiotemporal scales : The community structure of the Epibenthic fauna of the estuary and gulf of St. Lawrence, *Can. J. Fish Aquat. Sci.*, **47**, 1364～1377 (1990).

17) Gamito, S. and D. Raffaelli : The sensitivity of several ordination methods to sample replication in benthic surveys, *J. Exp. Mar. Biol. Ecol.*, **164**, 221～232 (1992).

18) 赤嶺達郎：底生生物の分布および種類組成を類別する統計的手法の検討，日水研報告，**33**，117～140（1982）.

19) 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）.

20) Williams, W.T. and W. Stephenson : The analysis of three-dimensional data (Sites × Species × Times) in marine ecology, *J. Exp. Mar. Biol. Ecol.*, **11**, 207～227 (1973).

21) 松宮義晴・和田恵子：時期×場所×種の採集データによる群集解析に関する一考察，西海区ブロック浅海開発会議 魚類研究会報，**5**，33～38（1987）.

22) Field, J.G., R.G. Green, F.A. de L. Andrade, E. Fresi, P. Gros, B.H. McAndle, M. Scardi, and D. Wartenberg : Numerical ecology : developments for studying the benthos, in "Developments in Numerical Ecology" (ed. by P. Legendre and L. Legendre), Springer-Verlag,

Berlin, 1987, pp. 485～494.
23) 品川汐夫：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）.
24) 品川汐夫・多部田修：数値実験の比較による Rsn 法の利点，日水誌，**64**，56～64（1998）.
25) 品川汐夫・多部田修：マクロベントスの調査データによる対応分析法と Rsn 法の比較，日水誌，**64**，418～426（1998）.
26) 品川汐夫・多部田修：河口域干潟における底生動物群集の経年変化についての Rsn 法による解析，日水誌，**64**，796～806（1998）.
27) 品川汐夫：動物の種－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，第10・11合併号，1～18（1992）.
28) Pielou, E.C.：The Interpretation of Ecological Data, 1st ed., John Wiley & Sons, New York, 1984, pp. 133～201.
29) Gower J.G.：Introduction to ordination techmiques, in "Developments in Numerical Ecology" (ed. by P. Legendre and L. Legendre), Springer-Verlag, Berlin, 1987, pp. 3～64.
30) 柳井晴夫：多変量データ解析法，初版，朝倉書店，東京，1991，pp. 31～51.
31) Gray, J.S.：Animal-sediment relationships, *Oceanogr. Mar. Biol. Ann. Rev.*, **12**, 223～261 (1974).
32) Zadeh, L.A.：Fuzzy sets, *Inf. Control*, **8**, 338～353 (1965).
33) 浅居喜代治，G.V. Negoita：あいまいシステム理論入門，第1版，オーム社，東京，1978，1～5.
34) Bezdek, J.C.：Some non-standard clustering algorithms, in "Developments in Numerical Ecology" (ed. by P. Legendre and L. Legendre), Springer-Verlag, Berlin, 1987, pp. 225～287.
35) 大隈 昇：ファジイ・クラスタリング，数理科学，**17**，34～41（1979）.
36) Roubens, M.：Fuzzy clustering algorithms and their cluster validity, *Eur. J. Ope. Res.*, **10**, 294～301 (1982).
37) 上田尚一：データ解析の方法，初版，朝倉書店，東京，1982，pp. 5～10.
38) McBratney, A.B. and A.W. Moore：Application of fuzzy sets to climatic classification, *Agricaltual and Forest Meteorology*, **35**, 165～185 (1985).

付　録

I．新座標づけ法の概要

2地点間の類似度 C_π' 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 **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 **y** とすると，**x** と **y** の類似度として，次の3つの因子を考えることができる．

1）方向の類似度　$\gamma = \mathbf{x}'\mathbf{y} / \|\mathbf{x}\|\|\mathbf{y}\|$

ただし，$'$ は行と列の転置を表すものとする．

2）総個体数の類似度　$\varepsilon_1 = 2\|\bar{\mathbf{x}}\|\|\bar{\mathbf{y}}\| / (\|\bar{\mathbf{x}}\|^2 + \|\bar{\mathbf{y}}\|^2)$

3）均等性の類似度　$\varepsilon_2 = 2d_x d_y / (d_x^2 + d_y^2)$

ただし，$d_x = \|\bar{\mathbf{x}}\| / \|\mathbf{x}\|$，$d_y = \|\bar{\mathbf{y}}\| / \|\mathbf{y}\|$

$\bar{\mathbf{x}}$，$\bar{\mathbf{y}}$ 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3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一側面を表しているから，xとyの類似度を，これら3つの因子の積として定義することができる．$\varepsilon_2\gamma$ は，木元の群集間類似度 C_π に一致するから，この3つの積を C_π' と定義する．

$$C_\pi' \equiv \varepsilon_1\varepsilon_2\gamma$$

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_π' 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 Rs の定義　2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，

$$\gamma \fallingdotseq \Sigma\, m_i^2\overline{\alpha\beta} / \|\mathbf{x}\|\|\mathbf{y}\|,$$　Σ は i についての総和とする．

そこで，α_i と β_i が独立な場合，

$$\gamma_{ind} \fallingdotseq \Sigma\, m_i^2\overline{\alpha}\ \overline{\beta} / \|\mathbf{x}\|\|\mathbf{y}\|$$

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

$$C_\pi' - C_\pi'{}_{ind} \fallingdotseq \varepsilon_1\varepsilon_2\gamma - \varepsilon_1\varepsilon_2\gamma_{ind} = \varepsilon_1\varepsilon_2\Delta$$

ただし，$\Delta = \Sigma\, m_i^2\,(\overline{\alpha\beta} - \overline{\alpha}\ \overline{\beta}) / \|\mathbf{x}\|\|\mathbf{y}\|$

となるから，2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 Rs を次式で定義する．

$\Delta > 0$　の場合：

$$Rs = (C_\pi' - C_\pi'{}_{ind}) / (1 - C_\pi'{}_{ind})$$
$$\fallingdotseq \varepsilon_1\varepsilon_2\Delta / \{1 - \varepsilon_1\varepsilon_2(\gamma - \Delta)\}$$

$\Delta < 0$　の場合：

$$Rs = (C_\pi' - C_\pi'{}_{ind}) / C_\pi'{}_{ind}$$
$$\fallingdotseq \Delta / (\gamma - \Delta)$$

地点間関係類似度 Rsn の定義　地点 i と j の相関指数 Rs_{ij} を成分とする対称行列［$\mathbf{Rsn}$］の，第 i 列と第 j 列のベクトル $\mathbf{R}_i$ と $\mathbf{R}_j$ 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 Rsn_{ij} を次式で定義する．

$$Rsn_{ij} \equiv \mathbf{R}_i'\mathbf{R}_j / \|\mathbf{R}_i\|\|\mathbf{R}_j\|$$

この操作により，Rsn_{ij} を成分とする対称行列［Rsn］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．

[Rsn]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，［Rsn］を最小二乗法の意味で最もよく近似する行列は，［Rsn］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，2つの地点の位置ベクトル間の余弦が，2地点の Rsn 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，［Rsn］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．

Ⅱ．ファジイ k-means 法の概要

Ⅰ．の方法によって地点サンプルと出現種を座標づけして空間に布置すると，多かれ少なかれいくつかの群が認識される．しかし，大抵の場合，その群ははっきりした境界が定め難く，多くの中間的な地点や出現種が存在する．このような地点と種相互間の関係を正確に把握するためには，ファジイクラスタリングが適している．ここでは，算法が比較的簡単な，ファジイ k-means 法を用いる．この方法は，通常のクラスター解析の一方法であるクラスター内分散最小化法を重みつき最小二乗法として扱い，その重みとして最適な各要素の所属率を，つぎの J_m の最小化によって求めるものである．

$$J_m = \sum_{i=1}^{n} \sum_{j=1}^{g} (u_{ji})^m \| \mathbf{x}_i - \mathbf{v}_j \|^2, \qquad 1 < m < \infty$$

ただし，$\mathbf{x}_i$, 要素 i の位置ベクトル（$i=1,2,\cdots,n$）

$\mathbf{v}_j$, クラスター j の重心座標（$j=1,2,\cdots,g$）

u_{ji} ,要素 i のクラスター j への所属率，$0 \leq u_{ji} \leq 1$，$\sum_{j=1}^{g} u_{ji} = 1$

g，クラスター数

m と g は，あらかじめ適当な数値を与えるものとする．多くの場合 m=2が採用されており，そのことによる不都合は報告されていない．

u_{ji} の局所最適解は，つぎの反復計算により得られる．

手順1：u_{ji} の初期値を適当に定める．

手順2：クラスター j の重心座標 $\mathbf{v}_j$ を次式により計算する．

$$\mathbf{v}_j = \frac{\sum_{i=1}^{n} (u_{ji})^m \mathbf{x}_i}{\sum_{i=1}^{n} (u_{ji})^m}, \qquad j=1,2,\cdots g$$

手順３：次式により u_{ji} を更新する．

$$u_{ji}=\frac{1}{\sum_{k=1}^{g}\left(\frac{\|\mathbf{x}_i-\mathbf{v}_j\|^2}{\|\mathbf{x}_i-\mathbf{v}_k\|^2}\right)^{\frac{1}{m-1}}},\quad i=1,2,\cdots,n,\qquad j=1,2,\cdots,g$$

手順４：旧 u_{ji} と新 u_{ji} の差があらかじめ定めた収束判定値（10^{-6}～10^{-7}程度）以下になれば計算を打ち切り，そうでなければ手順１に戻る．

Ⅲ．ファジイクラスタリング結果を評価するための指数

ファジイクラスタリングの結果を評価し，地点群と種群の対応関係を示すために，つぎのような指数を定義する．

1）種の平均地点群所属率と分布のあいまい度および地点のクラスタリングのあいまい度

種 p の，地点 i での出現個体数を x_{pi}，全地点での個体数の合計を N_p とし，種 p の地点群 j への平均地点群所属率 m_{pj} を次式で定義する．

$$m_{pj}=\frac{\sum_{i=1}^{n}x_{pi}u_{ji}}{N_p}$$

また，地点群 j への所属率を全地点について平均した値を U_j とする．

$$U_j=\frac{\sum_{i=1}^{n}u_{ji}}{n}$$

これらを用いて，種 p の，地点のクラスタリングに関する分布のあいまい度 Fz_p を次式で定義する．

$Fz_p=\min\{Fz_{p1},Fz_{p2},\cdots,Fz_{pg}\}$

ただし，$Fz_{pj}=\dfrac{1-m_{pj}}{1-U_j}$，　$j=1,2,\cdots,g$．min は最小値を意味する．

種 p の分布パターンが地点のクラスタリングとよく一致する場合には Fz_p は小さくなるであろうから，これを全種について平均すれば，地点のクラスタリングが種の分布と平均的に一致しない程度を表すことができる．これを地点のクラスタリングのあいまい度と定義する．

$$\text{地点のクラスタリングのあいまい度}=\frac{\sum_{p=1}^{S}Fz_p}{S},\quad S\text{は総種類数}.$$

2）種群と地点群の平均的関係および種群の分布のあいまい度

上に述べた m_{pj} および Fz_p を，種群別に種群への所属率の二乗を重みとして全種について平

均して，その種群と地点群 j の平均的関係を示す指数および種群の分布のあいまい度と定義する．すなわち，種 p の種群 i への所属率を v_{ip} とすれば，

$$\text{種群 } i \text{ と地点群 } j \text{ の平均的関係} = \frac{\sum_{p=1}^{S} v_{ip}^{2} m_{pj}}{\sum_{p=1}^{S} v_{ip}^{2}}$$

$$\text{種群 } i \text{ の分布のあいまい度} = \frac{\sum_{p=1}^{S} v_{ip}^{2} Fz_{p}}{\sum_{p=1}^{S} v_{ip}^{2}}$$

4）種群の個体数および種類数とその特化係数

各種の個体数を，種群への所属率の二乗を重みとして種群に配分して，全種について合計したものを種群の個体数と定義する．すなわち，種群の数を q として，地点 j における種群 i の個体数 z_{ij} を，

$$z_{ij} = \sum_{p=1}^{S} \left(\frac{v_{ip}^{2} x_{pj}}{\sum_{k=1}^{q} v_{kp}^{2}} \right)$$

と定義する．

z_{ij} の値は，種群および地点による個体数の多さに影響される．この影響を除き，地点の特徴を明瞭にするために，地点 j における種群 i の個体数の特化係数 c_{ij} を次式で定義する[37]．

$$c_{ij} = \frac{z_{ij} \div \sum_{k=1}^{n} z_{ik}}{\sum_{k=1}^{q} z_{kj} \div \sum_{k=1}^{q} \sum_{l=1}^{n} z_{kl}}$$

種類数についても同様に考えて，上の式で，x_{pj} の値を１（在）または０（不在）として計算した値を種群の種類数および種類数の特化係数と定義する．