

クラスター分析によるキノプロファイルデータのスクリーニング

中村 好宏¹・胡 洪管²・藤江 幸一³・馬場 康維^{4,5}

(受付 2004年3月3日; 改訂 2004年11月4日)

要 旨

活性汚泥内の微生物群集構造を特定する方法として、培養法によらないキノプロファイル法が注目されている。活性汚泥は細菌類や菌類を主な構成生物とし、原生動物や後生生物を従属生物群とした複合生物群であり、廃水処理系の浄化作用の中心的役割を果たしている。キノンは活性汚泥内の微生物により生成される化学物質で、キノ分子種の数、種類とその存在比、すなわちキノプロファイルの変化はとりもなおさず、微生物群集構造の変化に対応している。したがって、キノプロファイルの観察は活性汚泥内の微生物の構成を推定することにつながり、廃水処理系において微生物が分解する汚染物質の特定さらには、水質の評価に結びつくものとなる。胡らは、活性汚泥内のキノプロファイルを用いて汚泥内に存在するキノンの多様性の特徴づけを行っている。本稿では、この研究で用いられたデータを、キノプロファイルの季節性という観点から見直し、クラスタリングを行うことにより、先行研究とは異なる視点からデータの質の吟味を試みた結果を報告する。

キーワード：活性汚泥，キノプロファイル，季節変化，等質性に基づくクラスタリング。

1. はじめに

廃水の影響による河川の汚染，湖沼および内湾などの閉鎖系水域の富栄養化が問題になっており，効率が良い廃水処理系が必要とされている。廃水処理法には，物理化学的，生物学的方法が存在する。本論文では生物学的方法の1つである，活性汚泥による廃水処理に焦点を当てる。

活性汚泥とは，細菌類や菌類を主な構成生物とし，原生動物や小型の後生生物を従属生物群とした複合生物群で，水中の有機物を吸着，分解しながら呼吸，増殖を続ける一つの生態系である。水中の有機物を吸着，分解するというこの性質から，活性汚泥は下水処理や産業廃水処理などに広く使われている。しかし，活性汚泥内に存在している微生物の生態には，まだ解明されていないことが多い。より有用な廃水処理系の開発のためには，微生物群集構造の分析が

¹ 東京女子医科大学附属膠原病リウマチ痛風センター：〒162-0054 東京都新宿区河田町 10-22

² 清華大学 環境科学工程系：Beijing 100084, China

³ 豊橋技術科学大学 エコロジー工学系：〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

⁴ 統計数理研究所：〒106-8569 東京都港区南麻布 4-6-7

⁵ 総合研究大学院大学 複合科学研究科統計科学専攻

必要である。

Hu et al (2001) は、活性汚泥内のキノンの多様性に着目し、豊橋技術科学大学の下水処理施設で計測したキノンプロファイルを用いて、活性汚泥内に存在するキノンの特徴づけを行った。キノンは、呼吸鎖における電子伝達系物質の一つである。微生物が特定のキノンを有していることから、混合培養系におけるキノンの分子種の数、種類とその存在比、すなわちキノンプロファイルの変化は微生物群集構造の変化を表していることになる。活性汚泥を採取してその中に生息する微生物を数えることは手間がかかる。化学的にキノンを検出して微生物群の構造を特定することが可能なキノンプロファイル法は、簡便かつ有効なためその利用が広がりつつある手法である。

本論文は、クラスタリングによって、Hu et al (2001) のデータを再検討した結果の報告である。ここでは、キノンの季節変化に着目した。キノンプロファイルの季節変化を解析すれば、微生物群集構造の季節変化を把握することが可能になる。また、キノン種にこだわらず、季節変化をもつキノンとまたないキノンにグループ分けすることも、季節により存在量が変化する微生物の特定に有用な情報となる。したがって、キノンプロファイルの解析は微生物間の相互の関係の解明にも有用な情報となり得、廃水処理施設の処理能力の安定、向上に貢献できるであろう。ただし、データが 1 年分しかないため、モデルを用いた時系列解析の適用が困難である。そこで、クラスタリングによりデータの特徴を抽出し様々な可能性を示唆することとした。

本論文では、第 2 節にキノンプロファイルの説明を行い、第 3 節で本研究の目的と適用した手法の説明を行う。第 4 節で結果を示し、第 5 節で結果の考察と今後の展望を述べる。

2. キノンとキノンプロファイル

キノンは、呼吸鎖における電子伝達系物質の一つであり、ユビキノ、メナキノ、プラストキノ、ビタミン K₁ (フィロキノ) 等がある。一般に、ユビキノは好気性呼吸、メナキノは嫌気性呼吸、そして、プラストキノとビタミン K₁ は光合成にそれぞれ使われる。微生物に特定のキノンを有していることから、混合培養系におけるキノンの分子種の数、種類とその存在比、すなわちキノンプロファイルの変化は微生物群集構造の変化を表している。キノンプロファイル法は、微生物群集構造の変化を解析できる簡便、かつ有効な方法として、その利用が広がりつつある。

キノンプロファイルの特徴を表す指標として以下のものがある。

- ・キノンの数(n)、種類
- ・キノンの分子種の多様性(DQ)と分布の均一性(EQ)
- ・優占キノンの種とその存在比(fd)
- ・ユビキノとメナキノとのモル比(Q/MK)
- ・キノンの量

先行研究として、Hu et al (2001) が挙げられる。Hu らは、豊橋技術科学大学の下水処理施設で、21 項目のキノンに関する測定値と 7 項目の処理施設操作条件の計 28 項目を、1996 年 10 月から 1997 年 9 月までの 12 ヶ月間に渡り計測している(表 1, 表 2)。そこで得られたデータを用いて、優占キノンの種とその存在比、キノンの分子種の多様性が調べられている。その結果、いくつかの優占キノンの種が特定され、キノンの分子種の多様性が環境により変化することが判明した。本論文では、キノンの変化をより詳細に検討することを目的とし、このデータの更なる解析を行うこととする。

表 1. 活性汚泥における各種キノンの季節変化の測定値(単位: マイクロモル/g-粒状炭素).

月	10月	11月	12月	1月	2月	3月	4月	5月	6月	7月	8月	9月
Q-7	0	0	0.03	0.02	0.02	0	0	0	0	0	0	0
Q-8	0.29	0.31	0.46	0.39	0.2	0.23	0.27	0.29	0.2	0.19	0.18	0.31
Q-9	0.08	0.09	0.16	0.17	0.1	0.12	0.12	0.14	0.07	0.06	0.06	0.08
Q-10	0.13	0.19	0.3	0.26	0.14	0.24	0.16	0.19	0.11	0.12	0.11	0.15
Q-9(H2)	0	0.03	0	0	0	0	0	0	0	0	0	0
Q-10(H2)	0	0	0	0	0	0	0	0	0	0	0	0
VK1	0	0.01	0.01	0.01	0	0	0	0	0	0	0	0
MK-6	0.31	0.16	0.22	0.13	0.08	0.1	0.11	0.14	0.08	0.07	0.08	0.12
MK-7	0.4	0.29	0.47	0.25	0.14	0.18	0.2	0.23	0.13	0.11	0.15	0.25
MK-8	0.14	0.17	0.27	0.15	0.1	0.15	0.13	0.19	0.1	0.07	0.08	0.08
MK-9	0.07	0.07	0.06	0	0.06	0.07	0.06	0.07	0.04	0.02	0.04	0.06
MK-10	0.05	0.05	0	0.09	0.05	0.05	0.04	0.05	0.03	0.04	0.04	0.04
MK-11	0	0.01	0	0.05	0	0.01	0.01	0.02	0.01	0.01	0.01	0.01
MK-8(H2)	0.1	0.11	0.21	0.13	0.08	0.14	0.11	0.15	0.07	0.03	0.07	0.07
MK-9(H2)	0.04	0.03	0.13	0.05	0.03	0.05	0.03	0.04	0.02	0.02	0.02	0.02
MK-10(H2)	0.01	0.01	0.05	0.05	0.01	0.02	0.01	0.02	0.01	0.02	0.02	0.01
MK-11(H2)	0	0	0	0.01	0	0	0	0	0	0	0	0
MK-7(H4)	0	0	0	0	0	0	0	0	0	0.04	0.1	0.16
MK-8(H4)	0.04	0.03	0.06	0.04	0.02	0.03	0.02	0.02	0	0.02	0.02	0.04
MK-9(H4)	0.07	0.08	0.1	0.05	0.06	0.09	0.06	0.07	0.04	0.05	0.06	0.07
MK-10(H4)	0.12	0.09	0.15	0.06	0.05	0.11	0.07	0.1	0.06	0.07	0.08	0.09
合計	1.85	1.75	2.68	1.88	1.14	1.59	1.4	1.71	0.97	0.93	1.13	1.57

表 2. 処理施設の操業条件(単位: 水温(°C), 酸素濃度(mg/L), 流入水 BOD(mg/L), 処理水 BOD(mg/L), 汚泥沈降性(SV30), 水理滞留時間(d), BOD 負荷(kg-BOD/m³d)) (Hu et al. (2001)).

月	10月	11月	12月	1月	2月	3月	4月	5月	6月	7月	8月	9月
水温	24.5	21.9	19.8	17.0	17.0	17.9	18.8	21.9	23.9	26.0	27.0	27.5
酸素濃度	0.1	0.1	0.13	0.15	0.1	0.18	0.18	0.15	0.13	0.25	0.25	0.43
流入水 BOD	75	130	91	88	100	73	100	130	100	160	66	140
処理水 BOD	7.7	5.2	12	3	6.3	4	4.2	2.4	6.8	5.2	4	1.5
汚泥沈降性	95.6	96.8	95.8	96.2	92.5	95.8	95.5	95	95.5	96	82.3	80.5
水理滞留時間	0.64	0.64	0.69	0.85	0.75	0.85	0.67	0.63	0.63	0.67	0.83	0.8
BOD 負荷	0.12	0.2	0.13	0.1	0.13	0.09	0.15	0.21	0.16	0.24	0.08	0.18

3. 目的と手法

本論文の目的は、季節変化をもつキノンの特定である。廃水処理系に用いられる活性汚泥に関し、微生物群の構造の変化を把握できれば、処理施設の能力の向上や安定を目指す研究に対し、有用な情報を与えることが可能になる。

本研究を行うにあたり、季節変化の定義が重要な問題となる。前述の通り、季節変化とは、1年周期で現れる変動のことを指す。したがって、複数年のデータがなければ季節変化を特定することは困難である。1年分しかないデータによって、キノプロファイルがどのような変化をもてば季節による変化があるとみなせるかは、厳密に考えると難しい。ここでは、各キノンの反応パターンに基づいて、12ヶ月をクラスタリングし、連続する月が同じクラスターに所属しているような分割が得られたときに季節変化をもつと判断するものとした。この基準はキノンの変化が滑らかである場合に満たされるであろう。キノン量と季節に関連がないのであれば、

その値はランダムに動くか、別の要因に関連して変動するはずである。別の要因による季節変化をもつとすれば、このデータから季節変化を捉えることは難しい。しかし、上記の基準を満たすものが見つければ、周期的な変化をもつ可能性が高いと考えることができるであろう。

この基準で季節変化を特定するために、階層型クラスター分析法 (Everitt (1993)) を用いて月の系統樹を作成した。得られた系統樹が上記の条件を満たしていれば、キノンが季節変化をもつと考える。

しかしながら、階層型クラスター分析で得られた系統樹では、月、またはクラスター間の位置関係の把握が難しい。系統樹は、枝を中心に回転しても位相は変わらない。すなわち系統樹に表れる月間の距離は情報にならない。したがって、等質性に基づくクラスタリング (Nakamura and Baba (2000)) も適用することとした。この手法を用いることによって、クラスタリングとクラスター間の位置関係を同時に把握することが可能となり、クラスターの解釈が容易になるであろう (付録を参照)。

なお、階層型クラスター分析法は、S-Plus 4.5 を用いた。等質性に基づくクラスタリングは S-Plus 4.5 を用いてプログラミングしたものを利用した。

4. 結果

各キノンの季節変化の特徴を見るために折れ線グラフを描いたものが図 1 である。全体的に年末が高く夏季に低いという特徴が見られる。しかし、個々のキノンをみると、その特徴に該当しないものも多い。

はじめに、水温等の操業条件とキノン量の相関係数を求めたが、高い相関をもつ組み合わせはみられなかった。本来このデータは時系列データであるため、周期のずれを考慮したうえで相関係数を求めなければならない。同じカーブを描いていても周期がずれていると相関係数は高くないことがある。同様の理由で、操業条件による回帰分析も適用が難しい。したがって、本論文では操業条件は考慮せずに、キノン量のみを用いて解析を行うことにした。

第 3 節で述べた通り、クラスタリングによる季節変化の特定を行った。はじめに、すべてのキノンを用いて階層型クラスター分析法により月の系統樹を作成した。個体間の距離にはユークリッド距離、クラスターの結合法にはウォード法を用いた。結果を図 2 に示す。図 2 において、上の図では高さ 0.4、下の図では高さ 10 のあたりで枝を切ると、3 つのクラスターが生成される。この結果を見ると、どちらも季節変化に関する条件に当てはまっているようには解釈できない。しかしながら、全く外れているというわけでもない。この結果からは、ここで扱うキノンのすべてが季節変化をもつとは解釈できない。実際、折れ線グラフに示した通り、キノンには反応パターンが似たものと異なるものが混在する。したがって、より解釈が容易な結果を得るために、キノンのクラスタリングを行った。キノンには反応パターンが似たものと異なるものが混在しており、これをグループ分けして、キノンのグループ毎に月のクラスタリングを行えば、季節変化をもつキノンの集団をより把握しやすくなるであろう。

キノンのクラスタリングには、等質性に基づくクラスタリングを用いた。同手法は、各分類対象のプロファイルに基づいたクラスタリングであり、本研究で得られたデータの解析に有用であると考えられる。また、この手法を用いることにより、クラスター間の関係を視覚的に表現することが可能になり、結果の解釈が容易になると思われる。分類対象は 20 種のキノン (Q-10 (H2) は除く) とした。

ここでは 2 次元空間への射影を考えるものとする。各キノンを g 個のクラスターに分類した場合、生成された g 個のクラスターに便宜上番号をつけ、 $C_{g1}, C_{g2}, \dots, C_{gg}$ と表記するものとする。なお、各キノンの測定量は正規化した。これは各測定量の大きさではなく、プロフィー

ルのみで比較を行うためである。クラスタリングの結果と散布図を示す。ただし、等質性に基づくクラスタリングでは、 $g-1$ 以上の次元数の空間に射影しても、有用な情報となり得ない。したがって、クラスター数が2の場合の散布図は省略し、図3にはクラスター数を3~4としたクラスタリングの結果を記した。損失の変化を調べると、5グループ以上に分類しても損失はあまり減少しないことがわかる(図3の(d))。この結果より、キノンのクラスタリングにおけるクラスター数は3もしくは4とすることにした。

【2 グループへの分類】

C_{21} : {UQ-7, UQ-8, UQ-9, UQ-10, VK-1, K-6, MK-7, MK-8, MK-9, MK-8(H2), MK-9(H2),

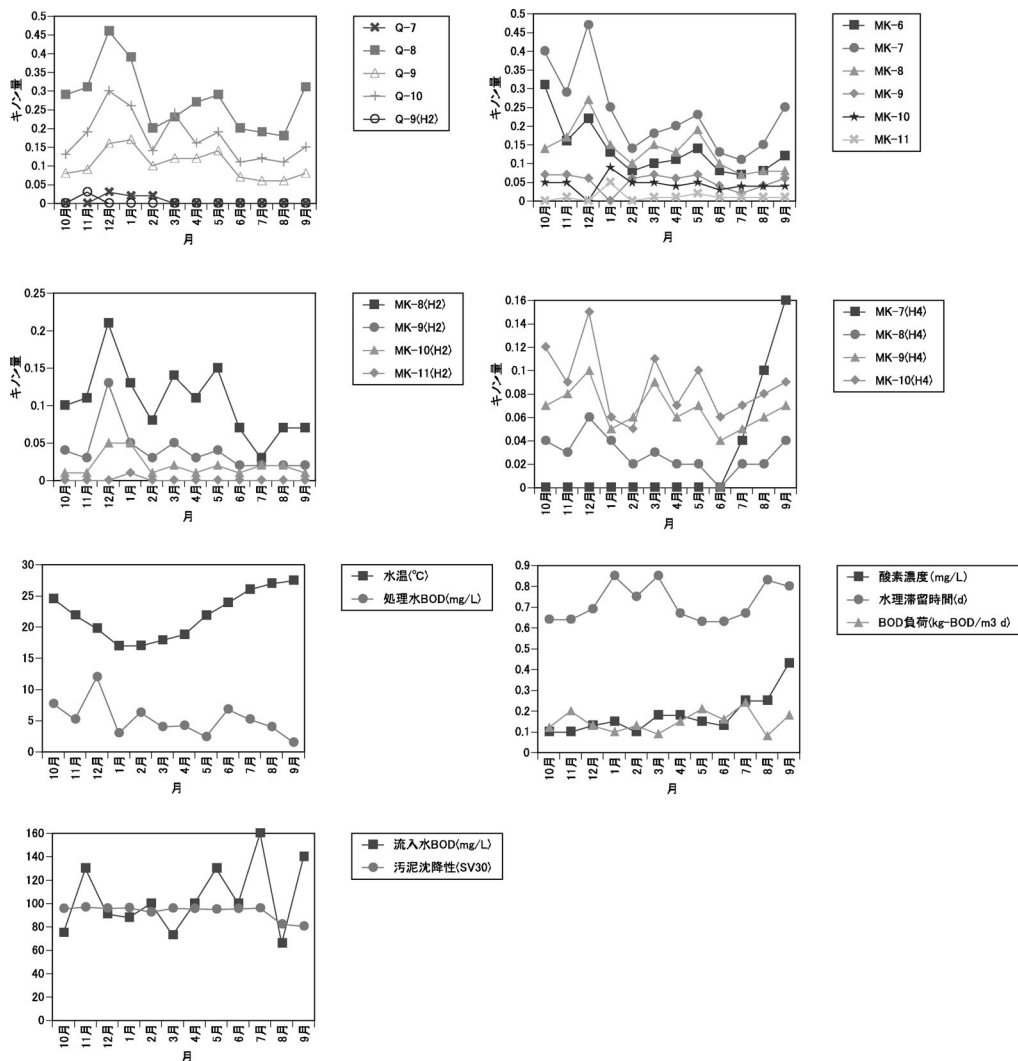


図1. キノン量と処理施設操作条件の折れ線グラフ。異なる反応パターンをもつキノンが混在していることがわかる。

MK-10(H2), MK-8(H4), K-9(H4), MK-10(H4)}

C_{22} : {UQ-9(H2), MK-10, MK-11, MK-11(H2), K-7(H4)}

【3 グループへの分類】

C_{31} : {UQ-7, UQ-8, UQ-9, UQ-10, VK-1, MK-7, MK-8, MK-8(H2), MK-9(H2), MK-10(H2), MK-8(H4)}

C_{32} : {UQ-9(H2), MK-6, MK-9, MK-7(H4), K-9(H4), MK-10(H4)}

C_{33} : {MK-10, MK-11, MK-11(H2)}

【4 グループへの分類】

C_{41} : {UQ-7, UQ-8, UQ-9, UQ-10, VK-1, MK-10(H2)}

C_{42} : {UQ-9(H2), MK-9, MK-7(H4)}

C_{43} : {MK-6, MK-7, MK-8, MK-8(H2), MK-9(H2), MK-8(H4), K-9(H4), MK-10(H4)}

C_{44} : {MK-10, MK-11, MK-11(H2)}

つぎに、得られたキノンのクラスター $C_{31} \sim C_{44}$ それぞれについて、再び月のクラスタリング

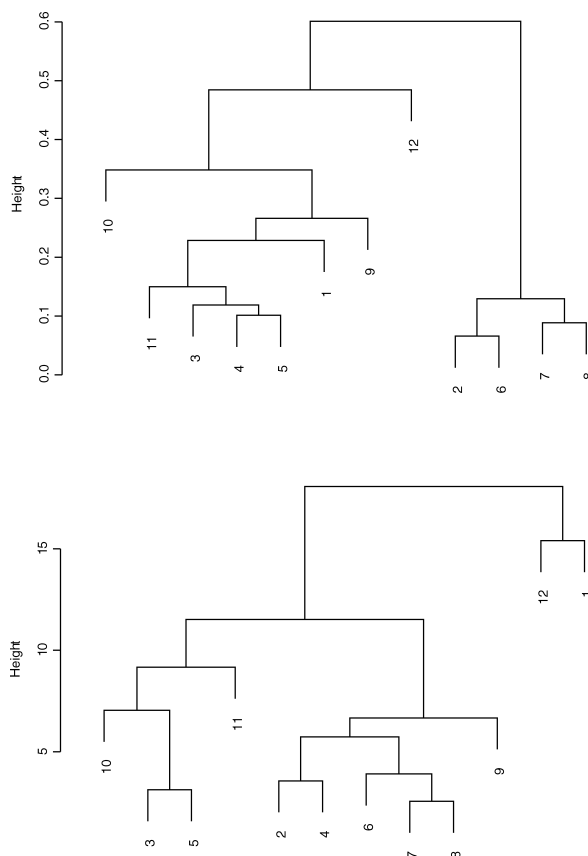


図 2. すべてのキノンを用いて月のクラスタリングを行った結果得られた系統樹。上が変数の正規化を行わなかったもの。下が行ったもの。どちらも隣り合った月が同じクラスターに配置されているとはいえない。しかし、全く外れているようにも見えない。

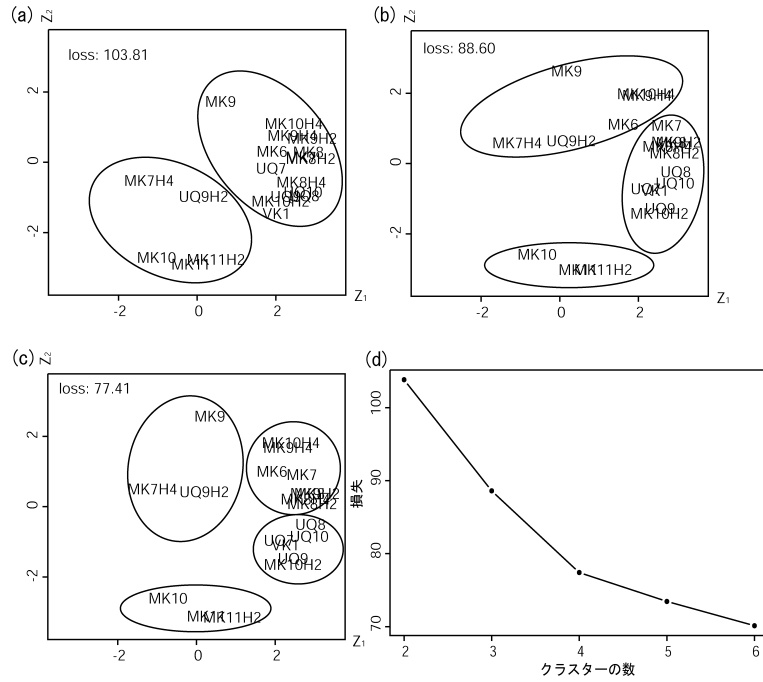


図 3. キノコのクラスタリングの結果と散布図：(a)クラスター数を 2 と仮定した場合；(b)クラスター数を 3 と仮定した場合；(c)クラスター数を 4 と仮定した場合；(d)クラスター数と損失の関係。この図は、クラスター数を 4 以上に増やしても意味があるクラスターは形成されるとは言えないことを示唆している。

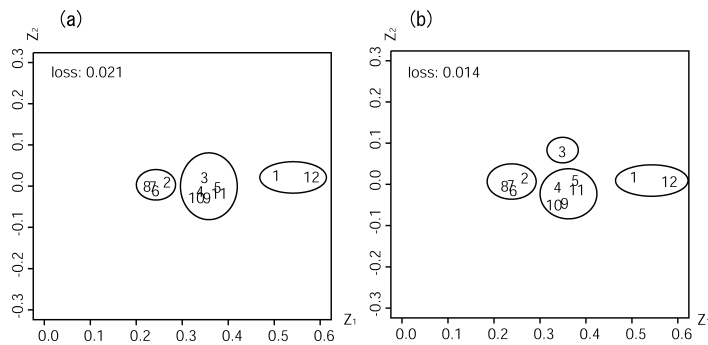


図 4. クラスタリングの結果と散布図：(a)クラスター C_{41} を用いた場合のクラスタリングの結果。クラスター数は 3 と仮定した。2 月の位置を除けば、季節変化があることを示唆した結果と解釈できる。(b)クラスター C_{41} を用いた場合のクラスタリングの結果。クラスター数は 4 と仮定した。2 月の位置は(a)と同様に不自然な配置である。更に、真ん中にある春と秋が混合した部分から 3 月だけが独立している。これは、春と秋の分類が上手く行われていないことを示している。

グを行った．ここでは12ヶ月を{春, 夏, 秋, 冬}に分類する場合と{春秋, 夏, 冬}に分かれる場合を考えて, クラスタ数 を 3 または 4 とした．結果は, クラスタ C_{41} に含まれるキノンの測定量を用いて月をクラスタリングした結果が本論文で定義した季節変化の条件に最も合うと思われる．クラスタ C_{41} を用いた結果を図4に示す．

しかし, 図4を見ると2月が6, 7, 8月と同じクラスタに配置されていることがわかる．この結果から判断すると, 2, 6, 7, 8月が含まれるクラスタが夏を表すクラスタとはいえず, 2月に連続している1月と3月は同じクラスタに配置されていない．従って, 季節変化の定義は完全に満足されておらず, 季節変化をもつキノンのグループは特定できなかったと結論付けることができる．

しかしながら, なぜ2月が6, 7, 8月と同じクラスタに所属するのかという疑問も生じる．図1において, C_{41} に属する UQ-7, UQ-8, UQ-9, UQ-10, VK-1, MK-10(H2) の時系列をみると, 1月から2月にかけて, キノン量が大きく減少している．2月のキノンは6, 7, 8月のそれと同等である．このデータのみで, 2月の測定値が正しく測定された値なのか, それとも低い確率で生じる誤差が大きい観測値だったのか, あるいは, なにか別の要因の影響があったのかは判断できない．しかし, 2月のデータを除いて考えると, C_{41} に含まれるキノンの季節変化をもつことを示唆する結果があらわれることもたしかである．これらの結果を考慮すると, このデータを用いて結論を導く前に, 再度データの質に関する吟味を行うことが必要であると思われる．

5. 今後の展望

本論文では, 1年間の月別キノンの時系列のパターンを用い, 類似性があるキノンをクラスタに分類した．さらに反応パターンが季節変化を持つと思われるキノンのクラスタを探索した．その結果, 季節変化を持つキノンのクラスタは特定できなかった．すなわち, 複数年にわたる計測を行い, 時系列解析等を用いて季節変化に関するエビデンスを示すことが重要であろう．2月のデータが単なる外れ値なのか, 構造的なものなのかはこのデータだけでは判断はできない．しかし, クラスタリングを行ったことによって, こういう問題点が洗い出されるという見地からは, 結論を導くものではなくとも問題の発見に統計的な手法が有効であったといえよう．いずれにせよ, 複数年のデータ収集が求められる．

今後は, 季節に伴う変化を持つキノンのその他のパターンをもつキノンについて, それぞれ変化の原因を調べ, 今回の解析結果と合わせて活性汚泥における微生物群集構造の変化を解明していきたい．

謝 辞

本論文をまとめるにあたって, 査読者および編集委員の方々より貴重なご意見をいただきました．ここに記して謝意を表します．

付録：等質性に基づくクラスタリング

等質性に基づくクラスタリングについて概観する．等質性とは, 複数変数間の関連性の指標であり, 損失関数により定義されている(Gifi (1990))． $n \times m$ のデータ行列を

$$X = (x_1, x_2, \dots, x_m)$$

とする．ここで, x_j は n 次元ベクトルである． m 個のベクトル x_1, x_2, \dots, x_m が, C_1, C_2, \dots, C_g で表される g 個のクラスタに分類されているものとする．クラスタ i に属するベクトルの

個数を m_i で表すと, $\sum_{i=1}^g m_i = m$ である.

以下の損失関数を定義する.

$$\sigma(C, X, Z, B) = \sum_{i=1}^g \sum_{j \in C_i} \|Z\beta_i - x_j\|^2.$$

ここで

$$Z = (z_1, z_2, \dots, z_k) \quad (n \times k)$$

は射影空間,

$$B = (\beta_1, \beta_2, \dots, \beta_g) \quad (k \times g)$$

は係数行列である.

C が固定されている場合, 上記の損失関数最小化は以下の固有方程式を解くことと同等になる.

$$XHX'z = \lambda z,$$

$$H = [h_{jj'}],$$

$$h_{jj'} = \begin{cases} m_i^{-1} & j \in C_i \text{ で } j' \in C_i \text{ のとき } (i = 1, \dots, g) \\ 0 & \text{それ以外.} \end{cases}$$

ただし, m_i はクラスター C_i の要素の個数である. 上記の固有方程式と, 変数の入れ替えを組み合わせたアルゴリズムにより, 最適な分割と射影空間 Z を得ることが出来る(中村(1999), 中村・馬場(1999), Nakamura and Baba(2000)).

参 考 文 献

- Everitt, B. S. (1993). *Cluster Analysis*, 3rd ed., Edward Arnold, New York.
- Gifi, A. (1990). *Nonlinear Multivariate Analysis*, John Wiley & Sons, Chichester.
- Hu, H. Y., Lim, B. R., Goto, N., Bhupathiraju, V. K. and Fujie, K. (2001). Characterization of microbial community in an activated sludge process treating domestic wastewater using quinone profiles, *Water Science and Technology*, 43(1), 99-106.
- 中村好宏(1999). 等質性に基づく多変量データの低次元空間への射影とその応用, 総合研究大学院大学数物科学研究科博士論文.
- 中村好宏, 馬場康織(1999). 損失関数最小化による多変量データの低次元空間への射影, 計算機統計学, 12(1), 7-21.
- Nakamura, T. and Baba, Y. (2000). An algorithm for clustering based on homogeneity, *The Tenth Japan Korea Joint Conference of Statistics Proceedings* (eds. Y. Tanaka and Y. Ahn), 63-68.

Screening of Quinone Profile Data Using Cluster Analysis

Takahiro Nakamura¹, Hong-Ying Hu², Koichi Fujie³ and Yasumasa Baba⁴

¹Institute of Rheumatology, Tokyo Women's Medical University

²Department of Environmental Science and Engineering, Tsinghua University

³Department of Ecological Engineering, Toyohasi University of Technology

⁴The Institute of Statistical Mathematics; The Graduate University for Advanced Studies

Activated sludge processes are the most widely used method for treating wastewater. However, only limited information is available on microbial community activities in activated sludges. In recent years, use of a technique using quinone profiles has increased as a simple and useful tool for analysis of microbial population dynamics in mixed cultures. Microbial respiratory quinones are components of the bacterial respiratory chain and play an important role in electron transfer during microbial respiration. Quinones exist in almost all bacteria, and in general, one species or genus of bacteria has only one dominant type of quinone. Thus, the quinone profile, which is usually represented as the mole fraction of each quinone type, should be specific for a microbial community. Changes in a microbial community in a mixed culture of microbes could effectively be quantified using quinone profiles. In this study, the technique of quinone profiles was applied to clarify seasonal change in the microbial community in an activated sludge process treating wastewater.

We tried to analyze the data of quinone profiles by descriptive multivariable analysis. Hierarchical cluster analysis and a clustering method based on homogeneity are applied in analyzing seasonal change of quinone profiles. In this study, clusters of quinones that have seasonal change could not be specified. However, it is suggested that the quality of the data needs to be determined.