

## (S5-22) 自己組織化マップと主成分分析・クラスター分析による

## 硝酸性窒素汚染地域の水質分類の比較

○天野弘基<sup>1</sup>・中川 啓<sup>1</sup>・朝倉 宏<sup>1</sup>・河村 明<sup>2</sup><sup>1</sup>長崎大学大学院・<sup>2</sup>首都大学東京

## 1. はじめに

長崎県島原市は、水道水源のほとんどを地下水に依存している一方、1975年以降、現島原市内の水道水源では、硝酸性窒素濃度が増加傾向にあり、現在では硝酸性および亜硝酸性窒素濃度の環境基準値を超過する水源もある。地下水の硝酸性窒素による汚染の現状は、著者らにより調査され、また地下水の水質特性はトリリニアダイアグラムやヘキサダイアグラムを描画することで評価された<sup>1)</sup>。

本研究では、同地域における水質調査データを元に、自己組織化マップ (SOM; Self-organizing Map) を作成し、水質特性パターンを抽出することで水質の分類をおこなった。SOMは、Kohonen<sup>2)</sup>により考案されたパターン分類手法であり、多次元データを2次元マップ上に表すことができるため、結果を視覚的に容易に判断でき、様々な分野<sup>3)6)</sup>で適用事例がある。SOMは、地下水水質の分類にも適用されている<sup>7)</sup>。また、本研究では、SOMを利用した水質分類結果と同地域における主成分分析 (PCA; Principal Component analysis) と階層的クラスター分析 (HCA; Hierarchical Cluster analysis) を用いた水質分類結果<sup>8)</sup>との比較をおこなった。

## 2. 研究対象地と使用データについて

研究対象地である島原市は長崎県南東部の島原半島北東に位置し、面積は82.8 km<sup>2</sup>である。島原市は、図1に示すように、半島の中心にそびえる雲仙普賢岳 (標高1,483 m) から東の有明海にかけてなだらかな斜面が広がる扇状地となっており、地質は玄武岩、安山岩そして凝灰角礫岩から主に構成される火山堆積物により形成されている。2012年における島原市の年間降水量は約2,540 mm、平均気温は約17°Cである。島原市では、北部に農業地帯が集中しており、北部における地下水および河川水では、硝酸性窒素濃度が高い傾向にある<sup>1)</sup>。

SOMの作成には、図1に示す民家の井戸 (RW, 12点)、三会小学校内に設置した観測孔 (O, 2点)、水道水源 (W, 20点)、湧水 (S, 4点)および河川 (R, 2点)の合計40地点における2011年8月から2013年11月の水質データ (Cl<sup>-</sup>、NO<sub>3</sub><sup>-</sup>、SO<sub>4</sub><sup>2-</sup>、HCO<sub>3</sub><sup>-</sup>、Na<sup>+</sup>、K<sup>+</sup>、Mg<sup>2+</sup>、Ca<sup>2+</sup> mg L<sup>-1</sup>) を使用した。データ総数は277である。SOM作成には、入力データ間のユークリッド距離が使われるため、SOM作成は非常にデータ前処理に敏感である。そのため、SOM作成の前に、全ての入力データに等しい、もしくは類似した重要性を持たせるために、データを標準化する必要がある。よって、本研究では入力データ (水質データ) を0から1の範囲に正規化した。

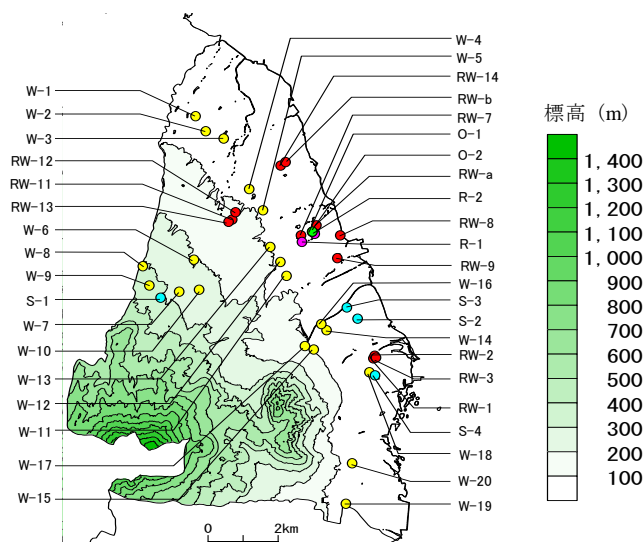


図1 研究対象地の概要

## 3. 自己組織化マップ

SOMは、入力データの類似性をマップ上に描画することのできるニューラルネットワークのひとつである。

Comparison of water quality classification by self-organizing map and principal components with cluster analysis  
in nitrate contaminated area

Hiroki Amano<sup>1</sup>, Kei Nakagawa<sup>1</sup>, Hiroshi Asakura<sup>1</sup> and Akira Kawamura<sup>2</sup>

(<sup>1</sup>Nagasaki University, <sup>2</sup>Tokyo Metropolitan University)

連絡先: 〒852-8521 長崎市文教町1-14 長崎大学大学院水産・環境科学総合研究科

TEL 095-819-2763 FAX 095-819-2763 E-mail: kei-naka@nagasaki-u.ac.jp

これにより、多次元のデータを2次元マップ上に分類し、可視化できる。入力ベクトルは、2011年8月から2013年11月の水質データ（8次元）により構成される。各ノードは、入力ベクトルと同次元の参照ベクトルを持ち、この参照ベクトルが入力ベクトルの特徴に漸近するように反復学習させる。最終的に、入力ベクトルは、最も類似した（ユークリッド距離が最も小さい）参照ベクトルに分類され、ノードマップ上での距離に近いほど互いに類似している。

SOMの全ノード数は、 $M = 5\sqrt{n}$ （ $M$ ：ノード数、 $n$ ：入力データ数）により決定され、縦列のノード数と横列のノード数は、入力データの共分散行列における固有値ベクトルの第1,2成分の比率から決定される<sup>9)</sup>。本研究では、縦12×横7（全ノード数84）のSOMが作成された。得られたSOM上で、参照ベクトルのクラスタリングをおこなう前に、最適クラスタ数を最小DBI（Davies-Bouldin Index）値により決定し、Ward法を用いてクラスタ分類をおこなった。

## 4. 結果と考察

### 4.1 自己組織化マップを利用した水質分類

作成された各イオンのSOMを図2に示し、最小DBI値により決定された5つのクラスタのSOM上の配置を図3に示す。図3におけるノード上の採水地点名は、その採水地点の最新データ（入力ベクトル）がそのノードに分類されてことを表している。図2と図3を比較することで、各クラスタの特徴を色調で視覚的につかむことができる。例えば、クラスタ1は、SOMの左上部に配置されていることから、すべてのイオンにおいて値が低く、SOM上の左下部に配置されたクラスタ5は、他のクラスタに比べCl<sup>-</sup>、NO<sub>3</sub><sup>-</sup>、K<sup>+</sup>の値が高いことがわかる。

表1に各クラスタに分類された全入力データ（各イオン濃度）の平均値を示す。図2と図3の比較から判断できたように、クラスタ1各イオン濃度は低い値を示し、クラスタ5のCl<sup>-</sup>、NO<sub>3</sub><sup>-</sup>そしてK<sup>+</sup>は、5つのクラスタの中で最も高い値を示した。クラスタ4とクラスタ5のNO<sub>3</sub><sup>-</sup>は、他のクラスタと比べはるかに高く、硝酸性窒素濃度に換算すると、どちらのクラスタも環境基準値の10 mg L<sup>-1</sup>を超過している。このことから、クラスタ4とクラスタ5は、硝酸性窒素による地下水汚染を示していると判断できる。クラスタ1、クラスタ2そしてクラスタ3のNO<sub>3</sub><sup>-</sup>は、人為的活動による影響の閾値である13 mg L<sup>-1</sup>以下であるため、これらのクラスタは島原市の本来の水質特性を示している。各クラスタのヘキサダイアグラムを描画したところ、クラスタ1、クラスタ2そしてクラスタ3はいずれもCa-HCO<sub>3</sub>型であったことから、これらのクラスタは溶存イオン量（表1）により区別される。また、クラスタ4とクラスタ5はどちらもCa-NO<sub>3</sub>型を示したことから、5つのクラスタはCa-HCO<sub>3</sub>型かCa-NO<sub>3</sub>型の2つに大きく分類される。

図4に採水地点40地点の最新データのトリリニアダイアグラムを示し、図5にクラスタで示した採水地点の分布を示す。図4によると各採水地点のサンプルは1サンプルを除き、アルカリ土類炭酸塩型かアルカリ土類非炭酸塩型に分類された。クラスタ1のみどちらの型にも分類された。クラスタ1に分類されたアルカリ土類非炭酸塩型のR-1やW-13は、SOMマップ上でクラスタ4に分類されたノードの隣であり（図3）、地理

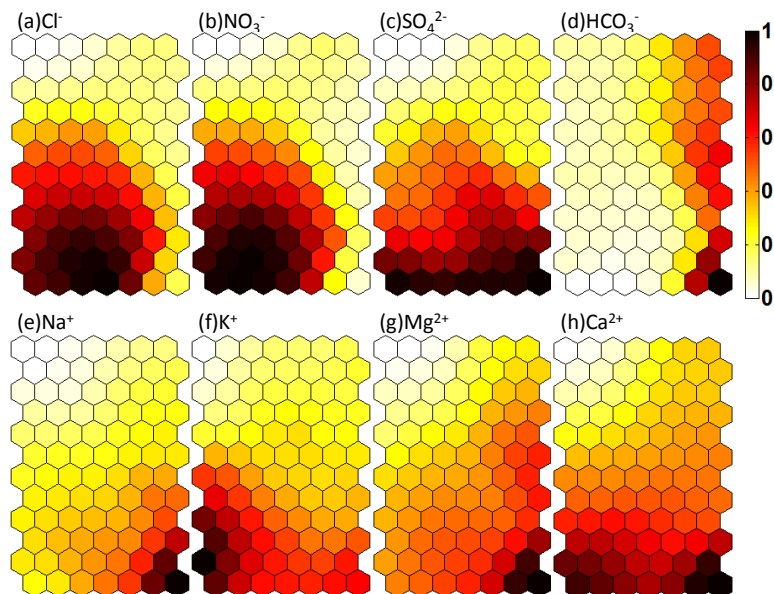


図2 主要溶存イオンのSOM

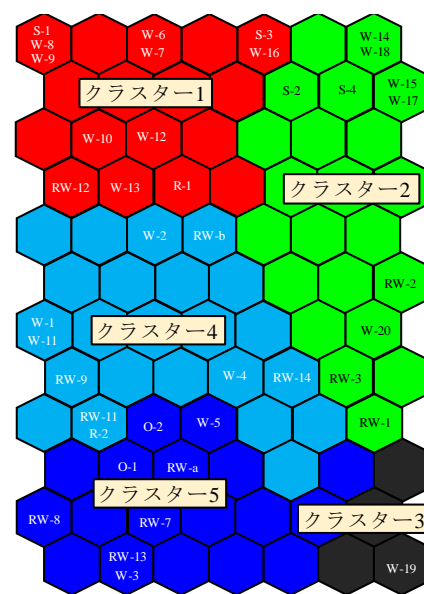


図3 SOMのクラスタ分類

的にもクラスター4に分類された採水地点の近くに位置する(図5)。これらは、クラスター4と類似した水質特性も有しており、硝酸汚染の影響を受けていると考えられる。

表1 SOMにより分類した各クラスターの主要溶存イオンの平均濃度

	Cl <sup>-</sup> (mg L <sup>-1</sup> )	NO <sub>3</sub> <sup>-</sup> (mg L <sup>-1</sup> )	SO <sub>4</sub> <sup>2-</sup> (mg L <sup>-1</sup> )	HCO <sub>3</sub> <sup>-</sup> (mg L <sup>-1</sup> )	Na <sup>+</sup> (mg L <sup>-1</sup> )	K <sup>+</sup> (mg L <sup>-1</sup> )	Mg <sup>2+</sup> (mg L <sup>-1</sup> )	Ca <sup>2+</sup> (mg L <sup>-1</sup> )
クラスター1	4.9	9.0	3.5	39.3	6.7	3.6	3.3	8.7
クラスター2	6.9	10.5	13.5	111.1	13.2	5.3	10.8	21.1
クラスター3	7.1	5.2	49.7	209.2	31.7	8.8	21.4	44.6
クラスター4	16.8	59.3	27.2	33.3	13.3	7.0	9.1	22.6
クラスター5	23.7	90.6	42.8	26.7	15.4	9.7	11.6	36.3

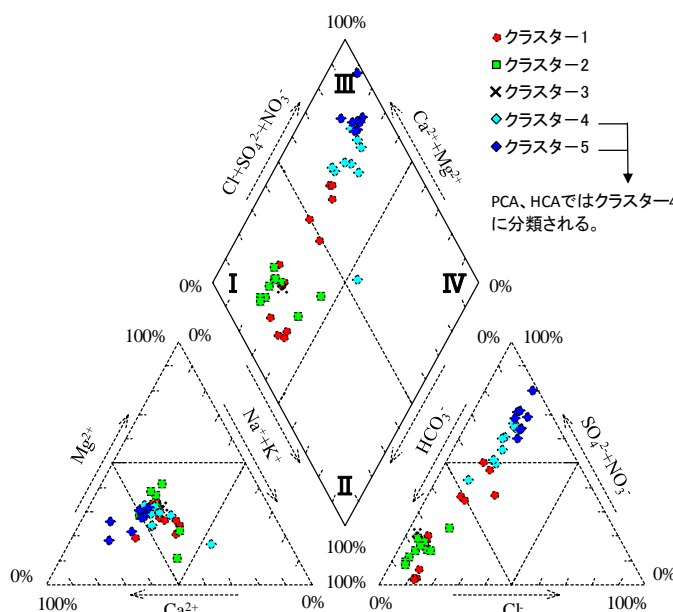


図4 各クラスターに分類した40地点のトリリニアダイアグラム

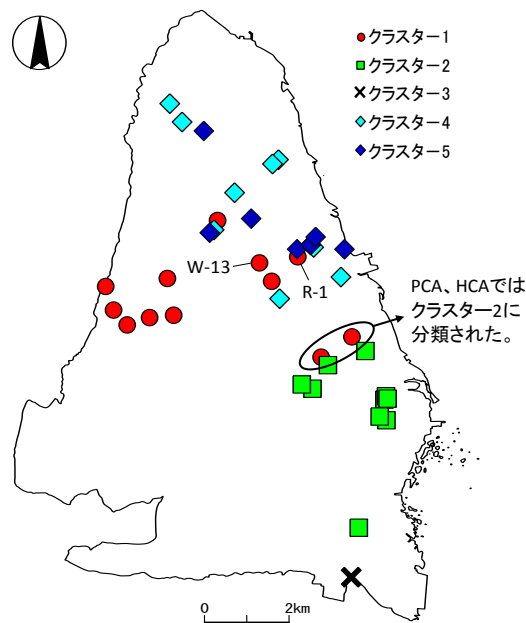


図5 クラスターで示した採水地点の分布

#### 4.2 PCAとHCAによる水質分類結果との比較

はじめに、SOMを利用した水質分類を方法1、PCAとHCAによる水質分類を方法2とする。どちらの方法においてもサンプルは同地点、同時期( $n=277$ )であり、クラスター分析はWard法を用いている。方法1と方法2の結果の違いは、まずクラスター数である。方法1においてクラスター数は、DBI値を用いて5と決定したが、方法2においてはデンドログラムをいくつかの位置で区切ることで最適と考えられるクラスター数を4と決定した。方法2における各クラスターの8種のイオンの平均濃度を表2に示す。表1と表2を比較すると、クラスター1、クラスター2そしてクラスター3の溶存イオン量(Cl<sup>-</sup>とNO<sub>3</sub><sup>-</sup>は除く)は非常に類似しており、クラスター間の濃度の違いは明白である。また、どちらの方法においてもNO<sub>3</sub><sup>-</sup>は13mg L<sup>-1</sup>を超えていない。このことから、方法1と方法2におけるこれら3つのクラスターは、それぞれ同様な意味を有するといえる。また、方法1のクラスター4とクラスター5、方法2のクラスター4はいずれも硝酸性窒素濃度が10mg L<sup>-1</sup>を超過しており、硝酸性窒素による地下水汚染という同様の意味を有する。つまり、方法1におけるクラスター4とクラスター5は、方法2においてはクラスター4にまとめられている。方法2においても各クラスターのヘキサダイアグラムはCa-HCO<sub>3</sub>型かCa-NO<sub>3</sub>型を示し、島原市の水質は大きく2つに分類され、クラスター1に分類された幾つかのサンプルは、トリリニアダイアグラム上で方法1の場合と同様にアルカリ土類非炭酸塩型を示した。

2つの方法による分類の違いは、各クラスターの分布で見られた(図5)。図5中の円で囲んだ採水地点(S-3、W-16)は方法2の最新データでは、クラスター2に分類された。このことから、方法2においては示唆されなかったが、方法1では地下水涵養域であると考えられるS-1やW-7~10が位置する森林部から、S-3やW-16が位置する沿岸部への地下水流れが示唆された。方法1と方法2ではクラスターの分類に違いはあるものの、クラスターの境界は明白であった。

表2 PCAとHCAにより分類した各クラスターの主要溶存イオンの平均濃度

	Cl <sup>-</sup> (mg L <sup>-1</sup> )	NO <sub>3</sub> <sup>-</sup> (mg L <sup>-1</sup> )	SO <sub>4</sub> <sup>2-</sup> (mg L <sup>-1</sup> )	HCO <sub>3</sub> <sup>-</sup> (mg L <sup>-1</sup> )	Na <sup>+</sup> (mg L <sup>-1</sup> )	K <sup>+</sup> (mg L <sup>-1</sup> )	Mg <sup>2+</sup> (mg L <sup>-1</sup> )	Ca <sup>2+</sup> (mg L <sup>-1</sup> )
クラスター1	5.3	11.7	3.8	37.7	7.0	3.7	3.5	9.2
クラスター2	6.7	9.9	13.6	113.0	13.2	5.3	10.9	21.0
クラスター3	7.1	5.2	49.7	209.2	31.7	8.8	21.4	44.6
クラスター4	20.0	73.4	34.7	31.0	14.3	8.3	10.4	28.9

## 5. おわりに

本研究では、硝酸性窒素による地下水汚染が問題となっている島原市の水質を自己組織化マップ (SOM) により分類した。島原市の水質は、DBI 値により 5 つクラスターに分類され、Ca-HCO<sub>3</sub> 型か Ca-NO<sub>3</sub> 型の 2 つに分類された。さらに、各クラスターはアルカリ土類炭酸塩型かアルカリ土類非炭酸塩型に分類されたが、クラスター1のみどちらにも分類され、島原市の本来の水質を示している一方で、硝酸汚染の影響を受けたものも含まれていた。SOMによる水質分類の結果は、主成分分析 (PCA) と階層的クラスター分析 (HCA) を用いた水質分類結果と非常に類似していたが、一部のクラスターの分布で違いが見られた。SOMを利用した各クラスターの分布から、PCAとHCAでは示されなかった方向への地下水流れが示唆された。

## 参考文献

- 1) 中川 啓, 天野弘基 (2013) : 島原市における地下水の水質特性について, 日本地下水学会 2013 年秋季講演会講演要旨, 198-199.
- 2) Kohonen, T. (1982) : Self-Organized Formation of Topologically Correct Feature Maps. *Biological cybernetics*, 43, 59-69.
- 3) 井芹慶彦, 水本真輝, 神野健二, 西山浩司 (2009) : 自己組織化マップを用いた日本全国月降水量分布の変動特性解析, *水文・水資源学会誌*, 22 (6), 466-478.
- 4) 横田いずみ, 井料隆太, 井芹慶彦, 広城吉成, 神野健二 (2009) : 自己組織化マップを用いた福岡市民の水に関するアンケート調査結果分析, *水工学論文集*, 53, 533-558.
- 5) 増野希陸, 原田昌佳, 平松和昭, 丸井 篤 (2012) : 自己組織化マップを利用した富栄養化貯水池の季節的な水質変動特性の定量的評価, *九大農芸誌*, 67 (1), 25-33.
- 6) 石原成幸, 河村明, 天口英雄, 高崎忠勝, 高橋泰之, 川合将文 (2014) : 自己組織化マップを用いた東北地方太平洋沖地震発生前後での東京の年間地下水変動特性, *土木学会論文集 B1*, 70 (4), I\_1129-I\_1134.
- 7) Nguyen, T.T., Kawamura, A., Tong, T.T., Nakagawa, N., Amaguchi, H. and Gilbuena, R. (2014) : SPATIAL CLASSIFICATION OF GROUNDWATER DATA IN THE RED RIVER DELTA, VIETNAM USING SELF-ORGANIZING MAPS. *Journal of Japan Society of Civil Engineers, Ser. B1(Hydraulic Engineering)*, 70(4), I\_241-I\_246.
- 8) 中川 啓, 天野弘基 (2014) : 多変量解析による島原市の地下水水質の分類, 日本地下水学会 2014 年春季講演会講演要旨, 2-3.
- 9) Hilario, L.G. and Ivan M.G. (2004) : Self-organizing map and clustering for wastewater treatment monitoring. *Engineering Application of Agricultural Intelligence*, 17, 215-225.

## 謝辞

本研究を実施するにあたり、地域の方々には井戸水のサンプリングにご協力いただきました。島原市水道局には、水道水源のサンプリングや各種関係資料の提供など、多大な協力を得ました。ここに記して謝意を表します。