

## 有害化学物質の水・底質環境における濃度予測

(株)エイトコンサルタント	正会員	○ 村口朝美	広島大学工学部	正会員	福島武彦
広島大学大学院	学生会員	大西英一	広島大学工学部	正会員	尾崎則篤
広島大学大学院	学生会員	嶋津治希	北海道大学工学部	正会員	東海明宏

国立環境研究所  
中杉修身

### 1. はじめに

化学物質の水(底)質濃度は、全国的に定点観測が行われているが、そのデータは膨大に存在する化学物質のほんの一部である。そのため、私達は安全かどうかはっきりしない状態で多くの化学物質を使用している。しかし、環境中の化学物質を全て測定するのは時間的、コスト的に不可能である。そこで化学物質のリスク評価をし、リスクの高い物質の判定が必要となる。本研究では化学物質のリスク評価のために、流域特性、化学物質の特性などから水(底)質中化学物質濃度を簡易に予測することが可能なモデルを作成することを目的とする。

### 2. 解析方法

対象物質は環境省刊行の「化学物質と環境」(1989～1997)より、28物質(64地点)を選んだ。これらは水質・底質共に4地点以上で検出されている。予測手法として、重回帰分析(以下重回帰モデル)と数量化分析I類(以下数量化モデル)を用いた。説明変数は表1に示した9つとした。流域は調査地点より上流側の全ての市町村を含む全流域と、調査地点に接する市町村のみを含む下流域の2つのパターンを考えた。流域に関する値は主に統計年鑑より、物質に関する値は文献やホームページより収集した。

### 3. 結果及び考察

重回帰モデル、数量化モデルを用いて、全物質・全地点(全入力)及び水域別(河川、海域、湖沼)で解析を行った。数量化モデルは説明変数のカタゴリー数を4つとし、各カタゴリーにサンプル数が等頻度になるようにした。その結果を図1(水質-全流域)及び表2に示す(ただし、数量化モデルの湖沼、水質-河川はサンプル数の少なさなどにより計算できなかった)。図1より、両モデルとも予測値は実測値から概ね±1オーダーの範囲に入っている。表3より、両モデルにおいて、(1)水質の方が底質よりR<sup>2</sup>は高い。SEは水質の方が全てにおいて小さい。これは、水質は堆積、吸着・脱着などの要素を持つ底質と違い、物質侵入時の影響を受けやすいためであると考えられる。(2)全流域と下流域では、若干下流域の方がR<sup>2</sup>は高いが明確な差はない。SEも同様である。(3)全入力と水域別では、重回帰モデルの水質-河川、湖沼を除いては全て水域別の方がR<sup>2</sup>は高い。SEも概ね同じ傾向である。これは、異なる流域特性を持つ水域を

表1 説明変数

流域に関するもの		物質に関するもの	
面積	ヘンリーランク	人口密度	logK <sub>ow</sub>
出荷額密度	生分解性	農地面積	生産量
	使用環境		

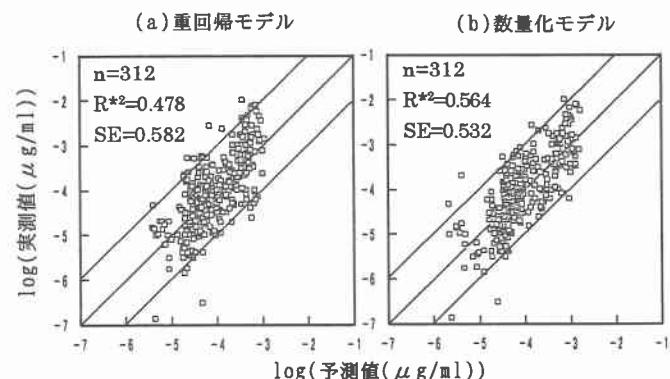


図1 水質-全流域の予測値-実測値

表2 重回帰モデル、数量化モデル結果

			n	R <sup>2</sup>	SE		R <sup>2</sup>	SE
水質	全流域	全入力	312	0.478	0.582	重回帰モデル	0.564	0.532
		河川	150	0.434	0.646		0.626	0.460
		海域	138	0.549	0.507		0.569	0.528
		湖沼	24	0.266	0.597		0.612	0.533
	下流域	全入力	312	0.482	0.580		0.596	0.477
		河川	150	0.449	0.637		0.448	0.650
		海域	138	0.513	0.526		0.478	0.607
		湖沼	24	0.307	0.579		0.496	0.645
底質	全流域	全入力	612	0.167	0.800	数量化モデル	0.464	0.641
		河川	245	0.203	0.751		0.479	0.606
		海域	311	0.181	0.824		0.511	0.635
		湖沼	56	0.269	0.715			
	下流域	全入力	612	0.169	0.799			
		河川	245	0.190	0.757			
		海域	311	0.175	0.826			
		湖沼	56	0.299	0.700			

n; サンプル数、R<sup>2</sup>; 自由度調整済み重相関係数、SE; 標準誤差

分けたことで説明変数を加えたような効果が起こり、それが良い方向に働いたと考えられる。(4)重回帰モデルと数量化モデルを比較すると、全てにおいて数量化の方が高い。特に底質では著しい向上が見られる。これは、数量化は非線形性を表現できるためであると考えられる。

次に、両モデルの検証を全物質・全地点(水質)を用いて行った。方法としては、全サンプルを無作為に半分に分け、片方をモデル構築に、片方をモデル検証に用いた。なお解析は10回繰り返して行った。その結果の平均値を表3に示す。数量化モデルの方が多いSEのばらつきが大きいが、両モデルともモデル構築と検証は同程度のSEを示している。よって、両モデルともモデル構築と同程度の精度(予測値が実測値から±1オーダーの範囲に入る)で予測が行えることが分かった。

重回帰モデルを用いて、物質別・地点別で解析を行った。その結果を図2に示す。図2より、予測値は実測値から±1オーダーの範囲に入っている。それぞれの物質(地点)について調べると、全物質・全地点での解析に比べてモデルへの適合度が良いことが分かった。

最後に、物質別・地点別(水質)についてモデルの検証を行った。方法は全物質・全地点の場合と同様である(表4)。これより、モデル検証の方が構築よりかなりSEが大きい。また、SEのばらつきも大きい。これは、10個程度のサンプルでモデルを構築したため、10個のサンプルに回帰式を合わせた形になり、モデルへの適合度は良いが予測精度が下がったと考えられる。そこで、安定した濃度予測を行うのに必要なサンプル数を調べるために、隅田川河口(水質)において物質数を増やしてモデルの検証を行った。追加した物質は過去に隅田川河口で検出され、その生産が現在も続いている8物質である。検証方法は、全サンプル29個のうちで構築用のサンプル(n)を10、15、20、25と増やしていく、残り(n')を検証に用いた。なお、解析は10回繰り返して行った。その結果を表5に示す。nが10のとき、モデル構築のSEは最も小さいにも関わらず、検証では最も大きい。また、SEのばらつきも最も大きい。それ以外では、モデル構築、検証共にn(n')が増える(減る)に従ってSEが小さくなっている。また、徐々にモデル構築と検証のSEが近づいている。よって、物質別・地点別モデルにおいてもサンプル数が15以上ある場合にはモデルの安定性が高くなり、精度の良い濃度予測を行えると考えられる。

#### 4. 結論

重回帰モデル、数量化モデルのようなモデル構築の容易な統計モデルを用いて、物性の異なる化学物質の環境中濃度を±1オーダーの範囲で予測できることが分かった。また、両モデルの使用方法として、重回帰モデルは、多くの物質・地点について解析を行う場合にも、ある一地点でいくつかの物質について解析を行う場合にも使用できる。数量化モデルは、多くの物質・地点について、特に非線形性を含むと考えられる場合に有効である。しかし、数量化モデルは予測を行う際に、モデル構築に用いた説明変数のデータ範囲に入らないデータの外挿ができないという短所を持っている。

表3 重回帰モデル、数量化モデルの検証(全物質・全地点)

		n	R <sup>2</sup>	SE	SEの幅
重回帰モデル	モデル構築	158.1	0.496	0.579±0.019	+0.029 -0.036
	モデル検証	153.9		0.602±0.021	+0.033 -0.025
数量化モデル	モデル構築	156.6	0.587	0.523±0.037	+0.070 -0.061
	モデル検証	155.4		0.583±0.042	+0.061 -0.080

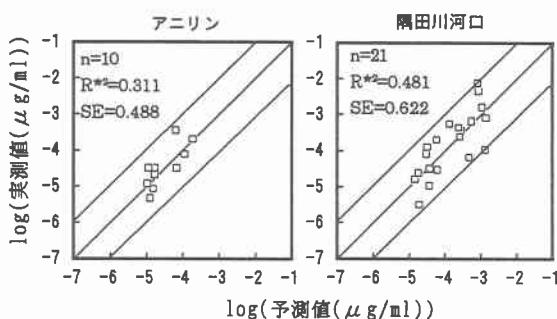


図2 物質別・地点別(水質)の予測値-実測値

表4 重回帰モデルの検証(物質別・地点別)

	物質・地点名	n	R <sup>2</sup>	SE	SEの幅
モデル構築	アニリン	7.5	0.226	0.427±0.155	+0.268 -0.290
	隅田川河口	10.5	0.600	0.478±0.157	+0.263 -0.276
モデル検証	アニリン	6.5		0.769±0.313	+0.602 -0.315
	隅田川河口	10.5		0.749±0.306	+0.733 -0.327

表5 サンプル数増加に伴う予測精度

	n(n')	R <sup>2</sup>	SE	SEの幅
構築	10	0.253	0.645±0.242	+0.415 -0.377
	15	0.085	0.771±0.106	+0.174 -0.232
検証	20	0.202	0.749±0.063	+0.107 -0.099
	25	0.208	0.728±0.031	+0.047 -0.045
	19		1.451±0.403	+0.741 -0.411
	14		0.811±0.132	+0.205 -0.221
	9		0.776±0.188	+0.444 -0.344
	4		0.707±0.213	+0.236 -0.379