

有害化学物質の水・底質濃度の 統計的予測に関する研究

福島武彦¹・大西英一²・嶋津治希³・尾崎則篤⁴・
村口朝美⁵・中杉修身⁶・東海明宏⁷

¹正会員 工博 広島大学大学院工学研究科教授 (〒739-8527 広島市鏡山1-4-1)

²学正会員 広島大学大学院工学研究科博士課程前期学生

³学生会員 広島大学大学院工学研究科博士課程後期学生

⁴正会員 工博 広島大学大学院工学研究科助手

⁵正会員 株式会社エイトコンサルタンツ (〒732-0055 広島市東区東蟹屋町15-3)

⁶正会員 工博 国立環境研究所化学物質環境リスク研究センター長 (〒305-8506 つくば市小野川16-2)

⁷正会員 工博 産業技術総合研究所化学物質リスク管理センター (〒305-8569 つくば市小野川16-1)

化学物質のリスク評価を目的に、物質特性、流域特性から水質、底質の有害化学物質濃度を予測する簡易なモデルの作成を目指した。まず、28物質、64地点における化学物質濃度、ならびにそれらに関する物質特性、流域特性を整理したデータベースを構築した。次に、重回帰分析、数量化理論Ⅰ類を当てはめた結果、両モデルとも全物質、全地点入力の場合、水質では実測値の±1オーダー、底質では±1.5オーダーの範囲内にほとんどすべてのサンプルが入ることがわかった。また、物質別・地点別モデル、説明変数の説明力、精度あるモデルの構築に必要なサンプル数などに関して議論した。

Key Words: toxic chemicals, forecasting model, multi-regression analysis, quantification theory type I, data base

1. はじめに

環境中には商業的に生産されたもの、非意図的に生成されたもの、あわせて一千万種類以上の化学物質が存在するといわれている。これら化学物質による環境汚染問題は、近年、複雑、多様化しており、その適切かつ効果的な管理・対策が早急に求められている。

こうした化学物質の管理・対策に当たっては、対象とする環境でリスクの高い化学物質をスクリーニングすることが必要となる。しかし、化学物質は種類が多いことから、すべてに対して濃度測定を行うことができないのが現状である。このため、対象とする環境の特性、化学物質の物性などから、環境中濃度を大まかに予測し、その結果をもとにリスクの高いことが予測される物質については濃度測定を行い、リスクを正確に算定し、管理・対策に役立てる、といった手順が考えられる。すなわち、本研究はどんな物質を調査すべきか、を決めるために多くの化学物質に適用可能な化学物質予測モデルの構築を目的とする。

水質・底質中の化学物質濃度を予測する簡易なモデルとしては既に以下のようなものがある。中杉は21物質、

全国12地点におけるデータを対象に、5つの化学物質特性、4つの流域特性を説明変数として変数選択型重回帰モデルを当てはめた¹⁾。水質、底質とも予測値は実測値の概ね1オーダー以内となることを示した。しかし、検出サンプル数が水質44、底質71と少なく、またモデルの検証は不検出データに対してのみ行っているだけで、実測検出データを用いての検証は行われていない。また、東海らは16物質、全国7地点でのデータ、ならびに6つの化学物質特性に対し、数量化分析Ⅱ類を適用し、予測値は実測値の1.2オーダーの範囲に入ることを報告した²⁾。しかし、地域域生に関する説明変数を入れていない、等の問題点を有している。上述のような統計モデルと異なり、大気、水、土壌などの相間の平衡モデルから化学物質濃度を予測したものとしては、盛岡・東海³⁾、Mackay & Hickie⁴⁾などがあるが、前者では3物質のみにしか検証を行っていない、後者ではPAHsの1地点、7物質のみしか対象としていない。すなわち、多くの物質、多くの地点に適用可能なモデルの構築という観点からの検証となっていない。

そこで本研究では、旧環境庁（現環境省）の発行した「化学物質と環境」⁵⁾をもとに、28物質、全国64地点

表-1 解析に用いた化学物質とその物性

化学物質名	解析に用いた地点数		ヘンリー定数 (25℃)	logKow	生分解率 (%)	生産量 (t/y)[1998]	使用環 境	用途	環境放 出相
	水質	底質							
アセトニトリル	13	11	3.45×10^{-5}	-0.34	80.0	4,750	B	d	固体
アクリルアミド	6	16	1.0×10^{-9}	-0.67	80.0	30,000	B	e	固体
アニリン	24	30	2.02×10^{-6}	0.90	74.0	213,490	A	e	固体
アントラキノン	3	13	2.35×10^{-8}	3.62	0	4,771	A	b	固体
ベンゾフルオランテン	0	44	5.84×10^{-7}	6.84	97.8	60	C	f	気体
ビスフェノールA	15	35	1.0×10^{-11}	3.32	68.5	258,310	B	b	固体
2-ブタノン	6	32	5.69×10^{-5}	0.29	0	124,701	B	d	液体
o-クロロアニリン	5	10	5.39×10^{-6}	1.92	4.0	500	A	b	固体
o-ジクロロベンゼン	14	14	1.92×10^{-3}	3.38	0	22,871	C	f	液体
m-ジクロロベンゼン	12	12	2.63×10^{-3}	3.60	0	22,871	A	e	液体
p-ジクロロベンゼン	19	19	2.41×10^{-3}	3.37	30.0	48,075	C	a	固体
N, N-ジメチルホルムアミド	6	5	7.39×10^{-8}	-1.01	0	36,000	B	a	液体
ジメチルスルホキシド	6	6	1.51×10^{-9}	-1.35	0	4,800	B	a	液体
1, 4-ジオキサン	22	22	4.8×10^{-6}	-0.27	62.4	4,500	B	e	液体
エタノールアミン	13	30	3.25×10^{-8}	-1.31	5.9	8,267	C	c	固体
イソプロチオラン	12	5	1.01×10^{-4}	3.32	80.0	2,513	C	a	液体
2-メチルピリジン	10	41	9.96×10^{-6}	1.11	8.6	1,667	C	b	液体
3-メチルピリジン	7	33	7.73×10^{-6}	1.2	5.9	1,667	C	b	液体
4-メチルピリジン	5	35	6.0×10^{-6}	1.22	0	1,667	C	b	液体
モルホリン	7	8	1.16×10^{-6}	-0.86	0	1,250	C	c	液体
ノニルフェノール	1	17	4.3×10^{-5}	3.28	2.0	20,000	C	c	液体
フェノール	30	45	3.33×10^{-7}	1.46	0	677,029	C	c	固体
ピレン	7	24	1.19×10^{-5}	4.88	80.0	1,949	C	f	気体
p-トルエンスルホンアミド	11	14	4.7×10^{-7}	0.84	80.0	500	B	d	固体
リン酸トリブチル	24	26	1.5×10^{-7}	4.00	4.0	245	B	d	液体
3, 5, 5-トリメチル-2-シクロヘキセン-1-オン	8	36	6.64×10^{-6}	1.70	80.0	4,500	C	a	液体
リン酸トリス(2-ブトキシエチル)	10	11	1.2×10^{-11}	3.75	0	2,384	B	d	固体
リン酸トリス(2-クロロエチル)	16	18	2.55×10^{-8}	1.44	0	2,384	B	d	固体

生分解率は28日間での数値。詳細は本文参照。

A; 閉鎖的使用(0.1), B; 半開放的使用(0.5), C; 開放的使用(1.0)

a; 農薬系, b; 工業薬品原料系, c; 家庭消耗品系, d; 一般工業製品添加物系, e; 染料・塗料系, f; 非意図的生成物

の水質・底質中の化学物質濃度データに加えて、それらの5種類の物質特性、4種類の流域特性を合わせたデータベースを構築して、重回帰分析と数量化理論Ⅰ類を当てはめた。次に、説明変数の説明力、予測精度、精度あるモデルの構築に必要なサンプル数、等に関して考察した。

2. 方法

(1) 解析用化学物質濃度データ

1989-1997年9年間の「化学物質と環境」⁹⁾をベースに、表-1に示す28物質(化学物質環境調査から24物質、水質・底質モニタリングから3物質、指定化学物質等検討調査から1物質)、全国64地点の水質・底質データベースを作成した。3物質(アントラキノン、ベンゾフルオランテン、ノニルフェノール; Shimazu et al.⁹⁾のデータベースから米国での測定データを除いたため)以外は、

水質・底質ともに4地点以上で検出されている。なお、各地点で測定値が多数ある場合は中央値を用い(中央値が検出限界未満となる場合は不検出データとした)、また多年度にわたる場合は嶋津ら⁷⁾と同じく各年度値の算術平均値とした。また、本研究の目的が化学物質濃度のオーダーを予測し得るモデルを構築することであり、さらに各物質の濃度の頻度分布を見ると、対数化により正規分布に近づくことから、以下の解析には濃度の常用対数値を用いた(図-1に対数化後の頻度分布を示す)。

なお、2.(3)に後述する地点別モデルでは、物質数を増やしての評価を行う必要があるため、上記の28物質に加えて、ピリジン、o-トルエンスルホンアミド、2-メチル-2,4-ペンタンジオール(化学物質環境調査)、ベンゾ[a]ピレン、BHT、m-ターフェニル(水質・底質モニタリング)、1,2-ジクロロエタン、1,2-ジクロロプロパン(指定化学物質等検討調査)の8物質を加えた。

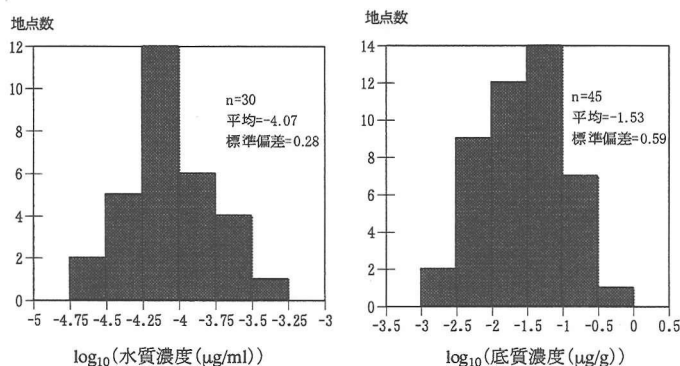


図-1 検出濃度のヒストグラム例（フェノール）

表-2 説明変数間の相関係数

	分子量	沸点	融点	蒸気圧	水溶解度	ヘンリー定数	logKow	生分解性	生産量
分子量	1								
沸点	0.621 **	1							
融点	0.223	0.624 **	1						
蒸気圧	-0.390 *	-0.464 *	-0.328	1					
水溶解度	-0.247	-0.104	-0.058	-0.031	1				
ヘンリー定数	-0.019	-0.118	-0.087	-0.102	-0.102	1			
logKow	0.731 **	0.698 **	0.419 *	-0.327	-0.380 *	0.285	1		
生分解性	-0.546 **	-0.418 *	-0.105	0.412 *	-0.083	-0.168	-0.369	1	
生産量	-0.146	-0.239	0.059	-0.013	-0.094	-0.058	-0.028	0.370	1

**; $p < 0.01$, *; $p < 0.05$ p: 有意確率

(2) 説明変数

化学物質の特性に関するものとして、分子量、沸点、融点、蒸気圧、水溶解度、ヘンリー定数、オクタノール／水分分配係数（logKow）、生分解性、生産量、使用環境、用途、環境放出相の12種類を選んで、データを収集した。しかし、表-2に示すように多くの変数間に高い相関関係が存在するので、相関係数ならびにそれぞれの変数の持つ意味を考慮して、ヘンリー定数、オクタノール／水分分配係数（logKow）、生分解性、生産量、使用環境（用途、環境放出相は質的データであるので、後述するように数量化理論Ⅰ類に対してのみ用いた）の5つを説明変数とした。以下にその変数の収集方法を説明する。

ヘンリー定数、logKowには、8)-12)に記載されている数値を並べ、より多くのデータベースで採用されている値を用いた。生分解性の大きさは、化学物質評価研究機構のデータ集¹³⁾にデータがあるものは、生分解が指数関数的に進むとして、その速度定数として与えた（ $C/C_0 = 10^{-at}$; C は1日後の濃度、 C_0 は初期濃度、 a は速度定数；底を10としたのは後のモデル化で濃度や説明変数の常用対数値を用いるため）。データがないものは予測システム¹⁴⁾を用いてその数値を推測した。すなわち、そのシステムによると「培養期間28日のBOD分解度が60%以上である

可能性が(a)高い、(b)低い、(c)非常に低い」という判定となるので、(a)の物質は28日間で80%、(b)は30%、(c)は0%の生分解率とした。なお表-1には、わかりやすくするため、期間を28日としてその間の生分解率を示した。生産量は14)から平成10年度のものを用いた。輸入量、輸出量がわかっている場合は生産量から足し、引きし、日本での消費量となるようにした。平成10年度の数値が得られない場合（8物質）は同じ統計の過去の資料（昭和58, 59, 61, 平成3年）から予測式をたて、推測した。なお、非意図的生成物であるビレンとベンゾ[k]フルオランテンは石油燃焼が起源と考え米国での発生予測量¹⁵⁾を基に、両国の原油消費量比を乗じて推測した。使用環境は、使用状態に関するキーワード（以下に例）を14)を用いて各物質に対してまず割り振った。個々の物質に対して複数のキーワードが割り振られる場合もあった。次に神奈川県先端産業立地指針を参考に、キーワードが農薬、洗剤、界面活性剤などの場合は開放的使用が行われるものとして使用環境を1.0、溶剤、触媒、洗浄剤などの場合は半開放的使用として0.5、中間物の場合は閉鎖的利用として0.1とした。キーワードが多数あり、使用環境の値が複数となる場合はその値が大きい方を用いた。

なお、数量化理論Ⅰ類の適用に際しては、化学物質

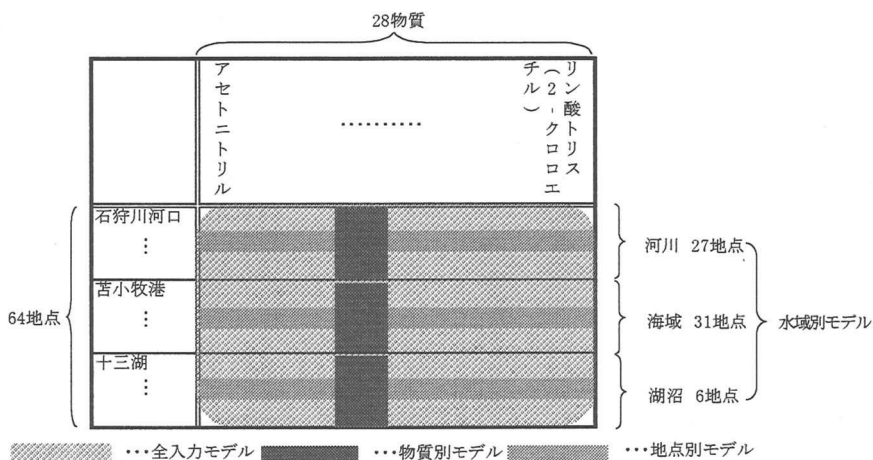


図-2 モデルの種類

特性として上述の5種類の説明変数を用いる場合と、さらに用途（農薬系、工業薬品系、家庭消耗品系、一般工業製品添加物系、染料・塗料系、非意図的生成物）と環境放出形態（気体、液体、固体）という2つのカテゴリー分類項目を加えた場合に対して、モデルの精度の比較を行った。用途と環境放出相の分類は14)をもとに行った。

次に流域特性を表現する変数として、面積、人口密度、出荷額密度、農地面積を選んだ。まず、各測定水域の流域となる市町村を地図から求め、それらの面積、人口、出荷額、農地面積の合計を算出し、面積以外は面積で除して密度とした。これを「全流域」データと呼ぶ。また、河川ではその測定地点が存在する市町村のみ、湖沼では周辺の市町村のみ、海域では測定地点の上流側の最近接市町村のみ、について算定を行ったものを「下流域」データと呼ぶ。なお、政令指定都市の場合には区単位で以上の計算を行った。市町村、区単位のデータは各都道府県の発行している平成9-11年度の統計年鑑から収集した。統計年鑑からデータが得られない場合は、面積は16)、人口は各県のホームページから値を求めた。

以上から、水質では312、底質では612の観測値とその説明変数をまとめたデータベースを構築した。なお、モデルの検証のため、検出限界未満のデータが得られた地点についてもその検出限界値を入れたデータベースを作成し、検出限界以上のデータで構築したモデルによってどのような予測値が得られるかを調べた。

(3) 統計解析方法

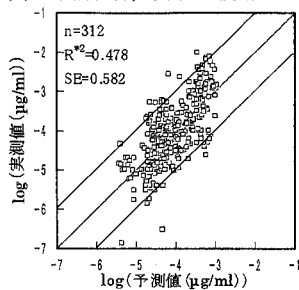
重回帰分析、数量化理論Ⅰ類（以降、数量化分析と略す）による分析には SPSS10.0 (SPSS 社) を用いた。両者とも、SPSS による変数選択は行わず、前節に挙げ

た説明変数を強制的に投入した。また、数量化分析に際しては、3.(1)に示すように説明変数のカテゴリー数を変えて適合度への影響を調べることも試したが、基本的にはカテゴリー数を4として解析を行った（使用環境のみ3カテゴリー）。この時、各カテゴリーに区分されるサンプル数がほぼ均等となるようにした。なお、サンプル数が少ない場合などには、説明変数のカテゴリー分類との関係で、数量化分析が行えないケースが存在した。

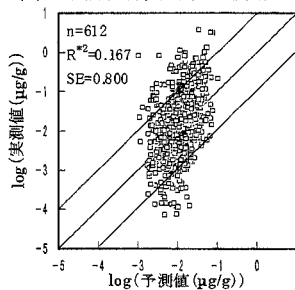
解析の単位としては、(1) すべての物質、地点（全入力モデル）、(2) 河川・湖沼・海域に分けてすべての物質（水域別モデル）、(3) 一つの物質で全地点、あるいは一つの地点で全物質（物質別・地点別モデル）、を対象とした3つのタイプの解析を行った（図-2）。なお(3)の解析は、サンプル数が少なく、また後述するように(1)と(2)の解析により「全流域」と「下流域」の結果の差が小さいことがわかったので、重回帰分析、「全流域」に対してのみ行った。(3)の物質別モデルでは3水域ごとに分けて解析し、サンプル数が6以上の物質を、地点別モデルではサンプル数が7以上の地点を対象とした。モデル精度の評価には、予測値と実測値の間の散布図に加えて、寄与率 (R^2)、自由度調整済みの寄与率 (R^2)、推定の標準誤差 (SE) を算定し、比較した。特に、モデル間の適合度の比較には R^2 を用いた。また各説明変数の説明力は、重回帰分析では偏回帰係数と有意確率、数量化分析ではカテゴリースコアとそのレンジにより評価した。

また、2.(1)にも述べたように水質・底質濃度は対数値とした。このため、説明変数についても対数値としたが、 $\log K_{ow}$ はもともと対数化された値、生分解性速度定数は指数部の係数なのでそのままとした。

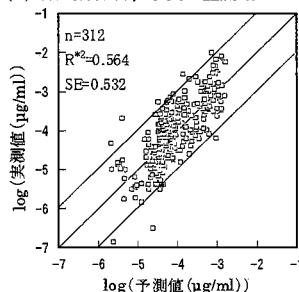
(1) 重回帰分析, 水質-全流域



(2) 重回帰分析, 底質-全流域



(3) 数量化分析, 水質-全流域



(4) 数量化分析, 底質-全流域

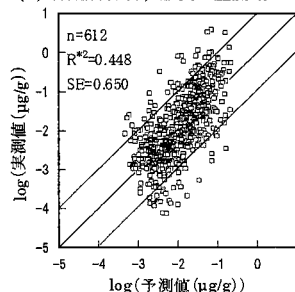


図-3 全入力モデルによる予測値 vs. 実測値

表-3 全入力モデル, 水域別モデルの結果一覧

			n	R ²	R ^{*2}	SE		R ²	R ^{*2}	SE	
重回帰分析	水質	全流域	全入力	312	0.493	0.478	0.582	数量化分析	0.577	0.564	0.532
			河川	150	0.468	0.434	0.646		0.650	0.626	0.460
			海域	138	0.578	0.549	0.507				
			湖沼	24	0.553	0.266	0.597				
		下流域	全入力	312	0.497	0.482	0.580				
			河川	150	0.482	0.449	0.637		0.635	0.612	0.533
			海域	138	0.545	0.513	0.526		0.623	0.596	0.477
			湖沼	24	0.578	0.307	0.579				
	底質	全流域	全入力	612	0.179	0.167	0.800		0.497	0.478	0.607
			河川	245	0.232	0.203	0.751		0.511	0.496	0.645
			海域	311	0.204	0.181	0.824				
			湖沼	56	0.389	0.269	0.715				
下流域		全入力	612	0.181	0.169	0.799	0.472	0.464			
		河川	245	0.220	0.190	0.757	0.498	0.479	0.606		
		海域	311	0.199	0.175	0.826	0.525	0.511	0.635		
		湖沼	56	0.414	0.299	0.700					

3. 結果と考察

(1) モデルの適合性

モデルの適合度（モデル間の比較は R^2 ）を各モデル、水質と底質、「全流域」と「下流域」で比較すると以下のようなものである（図-3、表-3）。

- ① 全入力モデルでは重回帰分析、数量化分析とも、水質では予測値は実測値の±1 オーダー、底質では±15 オーダーの範囲内にほとんどすべてのサンプルが入ったが（SE はそれぞれ 0.6, 0.8 程度）、底質に対する重回帰分析（図-3, (2)）では R^2 が 0.167 と低いため、

予測値の幅がかなり狭くなってしまっている。こうした±1 オーダー、あるいは±15 オーダーといった精度は、一般的に見れば悪いが、化学物質調査のスクリーニング、簡易な予測といった本研究の目的からみると十分なものと考えている。また、SE の大きさは既往の研究の結果とほぼ同じである¹²⁾。水質に比べて底質で適合度が低いのは、底質では堆積、吸着・脱着などの要素が複雑に絡んでいることと、化学物質濃度に大きく影響を及ぼす有機物濃度などの情報を説明変数に取り入れていないことが原因ではないかと考えている。また、重回帰分析と数量化分析

を比較すると数量化分析の方が適合度がよく、特に底質では差が大きい。これは、後者では様々な関数関係を表現しうることが原因ではないかと考えている。最後に、「全流域」と「下流域」を比較すると、「下流域」の方がわずかに適合度がよいもののほぼ同じであった。流域の上流より下流側の影響が大きいものの、差は大きくないことを意味している。

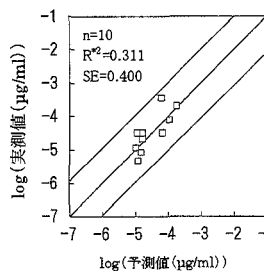
- (ii) 水域別モデルと全入力モデルを比較すると、大部分のケースで前者の方が適合度がよかった。これは、水域別とすることで説明変数が加わったような効果が生じたためではないかと考えている。なお、重回帰モデルの水質・河川、底質・海域は全入力モデルに比べて適合度が悪いが、また、水質では河川より海域の方が、底質では逆に海域より河川の方が適合度がよいが、これらの理由は不明である。

次に、物質別・地点別モデルに関しては以下のようなことがわかった。

- (iii) 全入力モデル、水域別モデルと比べ、大部分の物質、地域で SE は低下したが、 R^2 は増加したものの、減少したものの様々であり、サンプル数が少なくなったことが大きく影響を及ぼしていた（検出サンプル数が多い物質と地点の例；図-4、5）。SE が 0.5 未満となるのは、物質別モデルにおいては、水質・河川で 13 通り中 11 通り、水質・海域で 8 通り中 6 通り、底質・河川で 22 通り中 16 通り、底質・海域で 19 通り中 13 通りであった。また、地点別モデルにおいては、水質で 17 通り中 6 通り、底質では 42 通り中 7 通りであった。

- (iv) 地点別モデルで各物質の予測と実測の差の絶対値を誤差とし、ある物質が検出された全地点についてその平均を算出し、その物質の外れ程度の指標とした（水質のみ）。なお、絶対値をとる前の差の全地点平均値から、その物質の外れの傾向がわかる。また、外れ程度を絶対値をとる前の差の分散としなかったのは、その値の全地点平均値がゼロでないことと、また値そのものが外れ程度を表す絶対値の平均値の方がよいと考えたためである。同様に、物質別モデルで各地点の誤差を求め、ある地点で検出された全物質についてその平均を算出し、その地点の外れ程度の指標とした。この結果、半数以上の地点で平均値が 0.5（実測値と予測値の差が 1/2 オーダー）以上となる物質にはアクリルアミド（実測より高めに予測されることが多い）、1,4-ジオキサン（実測より小さめに予測されることが多い）、半数以上の地点で平均値が 0.25（実測値と予測値の差が 1/4 オーダー）以上となる地点としては新河岸川（実測より高めに予測されることが多い）、大阪湾、洞海湾（特徴なし）となった。すなわち、こうした物質、地点は外れ物質、外れ地点となる確率が高いことを意味して

(1) 水質-河川



(2) 水質-海域

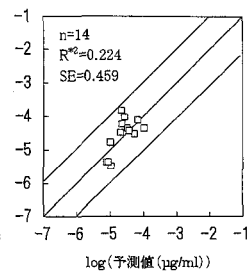
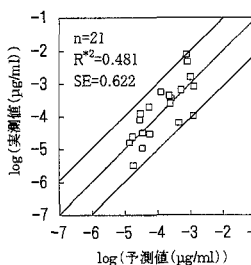


図-4 物質別モデル（アニリン）

(1) 隅田川河口（水質）



(2) 大和川河口（水質）

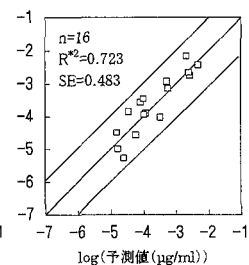


図-5 地点別モデル

表-4 外れ物質、外れ地点除外によるモデルの精度の変化（水域、全入力、「全流域」）

n	R^2	R^{*2}	SE	備考
312	0.493	0.478	0.582	全物質・全地点
284	0.503	0.487	0.547	物質除外
251	0.512	0.493	0.534	物質、地点除外

いる。このため、こうした物質や地点を全入力モデルから削除したところ、 R^2 は 0.015 増加し、SE は 0.048 減少した（表-4）。すなわち、何らかの理由で一定の傾向のある偏差を生じる物質、地点を選び出し、それらを除外してモデル化を行うことから、より精度あるモデルの構築が可能と考えられる。

ところで、数量化分析に際しては、カテゴリ数の変化と説明変数の追加に関して検討を行った。

- (v) 水質の全入力モデルに対して、各説明変数のカテゴリ数を 3 から 11 に変化させ、適合度の変化を調べた。図-6 には「全流域」の結果を示す。「全流域」の場合はカテゴリ数 8 で、「下流域」の場合は 9 で適合度が最大となった（カテゴリ数を 10、11 とすると計算が出来ない場合が生じた）。大部分の説明変数が元々、量的データであることから、こうした傾向を生じたものと考えられるが、サンプル数が少ない場合には計算が出来ないこともあるので注意しなければならない。

- (vi) 用途と環境放出相を説明変数に加えると、全入力モ

デル、「全流域」の水質では R^2 は 0.564 から 0.626 に増加し、底質では 0.448 から 0.609 と大幅に増加した。こうした変数はモデルへの寄与（カテゴリーレンジ）が大きいことが原因と考えられる。

(2) 偏回帰係数とカテゴリースコア

a) 入力モデルの偏回帰係数

理論的には、偏回帰係数の符号はそれぞれ水質、底質に対して、ヘンリー一定数では負、不明、 $\log K_{ow}$ は負、正、生分解性は負、負、生産量と使用環境は正、正となると予想される。また、流域に関する説明変数ではすべて正である。表-5に「全流域」に関して、水質、底質での偏回帰係数と有意確率をまとめるが、理論的なものと反対の符号になったものに記号を付けて示す。また、有意確率0.05以下のものに対して、 t 値の絶対値が大きい順に番号をつけてある。水質、底質ともに、 $\log K_{ow}$ 、使用環境といった説明変数の説明力が大きいことがわかるとともに、多くの有意な変数の偏回帰係数の符号、特に説明力の大きいものの符号は理論的に予想されるものと同じであることがわかる。符号が理論と逆になった説明変数は、説明力が小さく、また他の変数との相関関係の影響があらわれやすいものではないかと考えられる。

b) 物質・地点別モデルの偏回帰係数

水質の地点別モデルにおける $\log K_{ow}$ と使用環境の偏回帰係数の頻度分布を図-7に示す。 $\log K_{ow}$ は多くの地点で -0.5-0 に入り、使用環境では 0.5~1.5 に入る。すなわち、全入力モデルの結果とも、理論的な予測とも一致している。この他、ヘンリー一定数は 0.0 ± 0.2 、生産量は 0.0 ± 0.4 、生分解性は -0.2 ± 0.72 となっている。物質別モデルでもサンプル数が7以下の物質を除くと、人口密度 0.6 ± 1.6 、農地面積 1.2 ± 1.2 となるが、出荷額密度 -0.6 ± 1.17 、面積 -1.2 ± 1.3 となった。出荷額密度などは人口密度と相関が高いので、その影響で負となったものと考えられる。以上、すべてが理論通りとはならなかったが、主要な説明変数で予想された傾向であり、今回構築した重回帰モデルは化学物質動態の特性をある程度表現しているものと考えられる。しかし、今後、相平衡モデルなどを当てはめることから、そうした場合の分配係数と偏回帰係数の間の関係を解析する必要がある。

c) 数量化分析のカテゴリースコア、レンジ

図-8 は水質の全入力モデル（「全流域」）において、 $\log K_{ow}$ （レンジは 1.08）と使用環境（0.66）に対してカテゴリースコアを示す。両者とも理論的な傾向と一致している。他にカテゴリースコアの大きいものとしては生分解性（1.02）、生産量（0.88）があるが、それらでは理論的な傾向とは一致しなかった。底質ではヘンリー一定数（1.27）、 $\log K_{ow}$ （0.78）、生産量（0.59）の順となるが、この場合にはスコアの変化に一定の傾向が見られず、

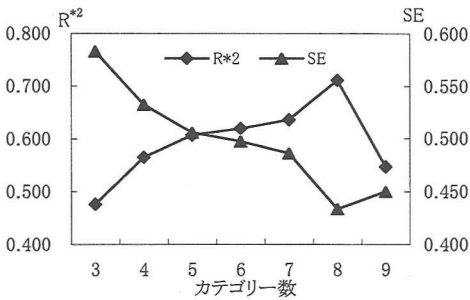


図-6 カテゴリー数のモデル精度への影響（水域＝「全流域」）

表-5 全入力モデルの偏回帰係数

(1) 水質－全流域

	パラメータ	偏回帰係数	t値	有意確率
	(定数)	-4.096	-8.436	0.000
	面積	-0.168 *	-1.342	0.180
	人口密度	0.242	1.666	0.097
	出荷額密度	0.034	0.277	0.782
	農地面積	0.173	1.567	0.118
	ヘンリー一定数	0.010 *	0.549	0.583
1	$\log K_{ow}$	-0.236	-11.495	0.000
	生分解性	1.574 *	0.642	0.522
	生産量	-0.020 *	-0.445	0.656
2	使用環境	0.935	7.779	0.000

(2) 底質－全流域

	パラメータ	偏回帰係数	t値	有意確率
	(定数)	-5.296	-12.194	0.000
	面積	0.280	2.818	0.005
5	人口密度	0.305	2.762	0.006
6	出荷額密度	0.211	2.083	0.038
7	農地面積	-0.124 *	-1.489	0.137
	ヘンリー一定数	0.026	1.362	0.174
1	$\log K_{ow}$	0.161	8.720	0.000
4	生分解性	6.897 *	2.838	0.005
2	生産量	0.146	3.982	0.000
3	使用環境	0.397	3.098	0.002

*: 理論的なものと反対の符号になったもの

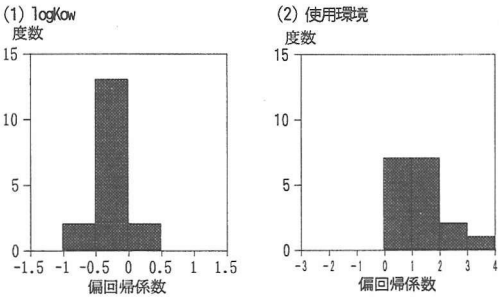


図-7 地点別モデルの偏回帰係数の頻度分布（水質）

理論通りではない。

(3) モデルの検証

a) 検出限界未満のデータを用いた検証（重回帰分析）

重回帰分析の全入力モデルに対して、検出限界未満のデータが得られたサンプルの説明変数を入力し、予測される値と検出限界値との関係を調べた。「全流域」では4割のサンプルで検出限界値以下、3/4以上のサンプルで検出限界値+0.5オーダーの、9割以上のサンプルで検出限界値+1オーダーの範囲内に入っており（図-9）、精度は悪くない。

b) 全入力モデルに対するデータ2分法（重回帰分析、数量化分析）

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ることから全入力モデルの予測性を調べた（水質の「全流域」）。サンプルの分割はケース選択の無作為抽出機能（SPSS）を用いて10回行った。重回帰モデルでは、モデル構築のSEは 0.579 ± 0.019 、モデル検証のSEは 0.602 ± 0.021 となり、数量化分析においても、それぞれ 0.523 ± 0.037 、 0.583 ± 0.042 と検証のSEの方が若干高いものの、ほとんど差がない。すなわち、 $n=150$ 程度になると、重回帰分析、数量化分析ともに、構築されたモデルのSEと同程度の精度で予測が可能であることを意味している。

c) 物質別・地点別モデルに対するデータ分割法（重回帰分析）

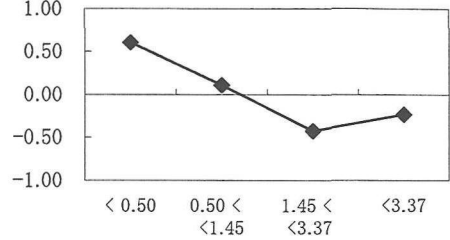
物質別・地点別モデルの中ではサンプル数の多いアニリン、フェノール、隅田川河口、大和川河口に対して、データを2分してモデル構築とモデル検証を行った。この結果、どのケースとも構築のSEと比べ検証のSEが大きく増加した（表-6）。すなわち、データが少ないために、構築されたモデルが構築データの特性に強く規定され、検証データをうまく表現できていないのではないかと考えられる。

このため、化学物質の種類を増加させ（地点別モデルのみ）、モデル構築のサンプル数を増加させたときのモデル検証のSEの変化を調べた。隅田川河口では、構築用サンプル数を10、15、20、25と増加させるにつれ、検証のSEは低下し、その数が20程度になると構築のSEとほぼ同じとなった（図-10）。すなわち、モデル構築のサンプル数を20以上にすればモデル構築の精度と同程度の精度で予測が可能であることを意味している。大和川河口でも同様なことがわかった。

4. モデルの比較

数量化モデルは重回帰分析と比べて、様々な関数関係を表現できる、質的データを取り入れられるなどの長所を有する一方、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難しい、解析に必要なサンプル数が多い、といった一般的な短所

(1) logKow



(2) 使用環境

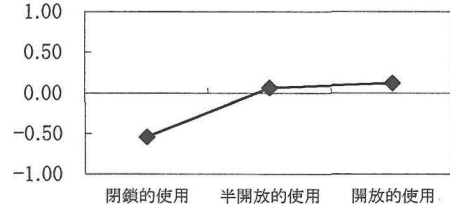


図-8 カテゴリースコアの例（水質—「全流域」）

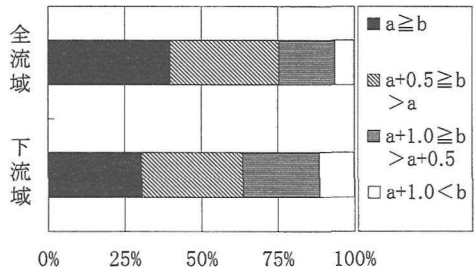


図-9 検出限界未満のデータを利用した重回帰分析の検証（水質—全入力）（a：検出限界値，b：予測値）

表-6 物質別・地点別モデルの検証結果（10回の平均値）

	物質・地点名	n	SE
モデル構築	アニリン	7.5	0.427 ± 0.155
	フェノール	8.4	0.273 ± 0.191
	隅田川河口	10.5	0.478 ± 0.157
	大和川河口	8.1	0.528 ± 0.188
モデル検証	アニリン	6.5	0.769 ± 0.313
	フェノール	8.6	0.442 ± 0.369
	隅田川河口	10.5	0.749 ± 0.306
	大和川河口	7.9	0.899 ± 0.390

を有している。今回の全入力モデル、すなわちサンプル数が150（検証の際のサンプル数）を超える場合は、水質、特に底質で、数量化分析の方が重回帰分析より精度が高いことがわかり、予測モデルとしても十分役に立ちそうである。これに対して重回帰分析は、全入力モデルの水質では数量化分析と同程度の精度を持つとともに、

物質別・地点別モデルにおいてもサンプル数を 20 以上とすればかなりの精度の予測が可能と考えられた。

全入力モデルと物質別・地点別モデルの使い分けに関しては、今研究のサンプル数では全入力モデルの方が物質別・地点別モデルより検証のSEが小さく、精度がよいことがわかった。しかし、後者の方が流域特性あるいは物質特性の影響を受けないことから、サンプル数がさらに増えればより精度のあるモデルになってゆくことが期待される。今後、各地点、各物質に対するモデルを増やし、また各モデルでサンプル数を増加させることから、重要な説明因子の解明と物理モデルとの対応を図ってゆく必要がある。

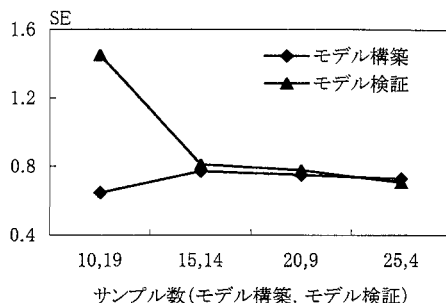


図-10 モデル構築のサンプル数増加に伴うモデル精度の変化 (隅田川河口)

5. おわりに

水質・底質中化学物質濃度予測の手法として、重回帰分析と数量化分析を試した。全入力モデルでは、両分析ともに水質では実測値の ± 1 オーダー、底質では ± 1.5 オーダーの範囲内にほとんどすべてのサンプルが入ることがわかった。また、サンプル数ある程度以上、大きくするとモデル構築の精度と同じくらいの予測精度が得られることを示した。今後、重要な説明変数ではあるが、今回収集できなかった各地域での排出量、底質中の有機物濃度などを取り込み、精度の向上を図る必要がある。また、物質別・地域別モデルの数を増加させ、causalityを有する物理モデルとの対応を図ってゆくことが重要と考える。

参考文献

- 1) 中杉修身：化学物質の運命の統計予測モデル、化学物質環境運命予測手法開発調査、日本環境協会、pp.1-24, 1985.
- 2) 東海明宏、長屋圭治、湯浅晶：数量化分析を用いた環境中化学物質濃度区間の推定、環境化学、6, pp. 503-510, 1996.
- 3) 盛岡通、東海明宏：化学物質の環境リスク評価のための運命予測モデルの適用、衛生工学研究論文集、23, pp.115-124, 1987.
- 4) Mackay, D. and Hickie, B.: Mass balance model of source apportionment, transport and fate of PAHs in Lac Saint Louis, Quebec. *Chemosphere*, 41, pp. 681-692, 2000.
- 5) 環境庁環境保健部環境保全課：化学物質と環境、1989-1998.
- 6) Shimazu, H., Ohnishi, E., Ozaki, N., Fukushima, T. and Nakasugi, O.: A model for predicting sediment-water partition of toxic chemicals in aquatic environments, *ASIAN WATERQUAL 2001*. (in press)
- 7) 嶋津治希、大西英一、尾崎則寛、福島武彦、中杉修身：化学物質の水中濃度と底質中濃度の関係、水環境学会誌、23, pp. 786-794, 2000.
- 8) <http://escsytres.com/interkow/physdemo.htm>
- 9) <http://www.nihs.go.jp/ICSC>
- 10) <http://www.k-erc.pref.kanagawa.jp>
- 11) 環境庁環境化学研究会：環境化学物質要覧、丸善、1988.
- 12) Lide, D. R.: *Handbook of chemistry and physics 75th edition*, 1994-1995.
- 13) <http://www.oetj.or.jp>
- 14) 化学工業日報社：13700の化学商品、2000.
- 15) U.S. Environmental Protection Agency: An exposure and risk assessment for benzo(a)pyrene and other polycyclic aromatic hydrocarbons, Vol. III, PB85-222578, Vol. IV, PB85-222586, 1985.
- 16) <http://www.gis.go.jp>

STATISTICAL FORECAST OF THE CONCENTRATION OF TOXIC CHEMICALS IN WATER AND SEDIMENT

Takehiko FUKUSHIMA, Eiichi ONISHI, Haruki SHIMADU, Noriatsu OZAKI, Asami MURAGUCHI, Osami NAKASUGI and Akihiro TOKAI

The preparation of the models forecasting the concentrations of toxic chemicals in water and sediment by the use of their chemical and watershed properties was examined to apply the models to the risk-assessment of them. The database on 28 chemicals and 64 stations was constructed and applied for calibrating and validating the statistical models i.e. multi-regression analysis and quantification theory type I. In the both models and irrespective of analysis type, there observed within one order of magnitude difference in the case of water and 1.5 orders of magnitude difference in the case of sediment between the forecasted and the observed values for almost all the data. The models on an each chemical basis or on an each station basis, the dominant input variables, and the number of samples for constructing the model having sufficient accuracy were also investigated and discussed.