

確率統計的手法による大気汚染予測

EXPERIENCE IN AIR POLLUTION PREDICTION BY STATISTICAL
AND STOCHASTIC PREDICTION TECHNIQUES

日野 幹雄*・吉川 信二郎**・栗原 達雄***

By Mikio HINO, Shinjiro YOSHIKAWA and Tatsuo KURIHARA

1. 序 論

大気および水質の汚染については、環境基準が設定されこれを守るべく汚染排出物の浄化除去の努力が続けられている。公共団体や工場のこうした努力により、汚染状況は改善の方向に向ってはいるが、いまなお環境基準をオーバーする汚染濃度が計測される地域も少なくないといわれる（特に、 NO_x について）。環境濃度が基準値を越えることが、直ちに人体および動植物の生命の危険を意味するものではないが、やはり設定された基準を守るべきであり、それを越えると予想される場合には警報を出すなり、工場の操業度を落すよう要請が行われなくてはならない。こうした立場から、環境汚染度および気象条件を刻々モニターし、数時間先の汚染濃度をオンライン的に予測し、必要ならば工場操業を規制しようという APMS (Air Pollution Monitor & Control System) プロジェクトが機械振興協会により立案され（昭和 46 年）、昭和 50 年度には実際に茨城県鹿島工業地区において実施に移されている。これまでは、環境のモニタリングは熱心に行われてきているが、それを活用して予測が実際に行われている例はほとんどなく、収集されるデータが十分生かされていないように思う。

著者の一人は、これまでいくつかの汚染予測法について提案し検討を重ねてきた^{1)~4)}。すなわち、日周期性を考慮した回帰予測、拡散システム方程式の係数同定にカルマン・フィルター理論を応用する予測法、これらに因子分析を組み合わせる FRAC 法である。今回、これらの予測方法を上述の APMS の一部として実際に試用することができたので、その結果について報告したい。

2. 統計的モデルの物理的意味づけ

回帰分析・因子分析等の多変量解析の理論は、心理学・医学・農学のような元来対象の変化を支配する普遍的な基礎法則のない学問分野において有力な解析方法として発達してきた。その後、この方法は気象学における天気予報のように、本来物理法則の支配を受ける現象であるにもかかわらず、要素間の相互関係の強い複雑な現象にも応用されるようになってきた。大気汚染もまた気象という自然環境の中における人間の社会活動にもとづく汚染物の放出という入力により発生する物理現象であるが、大気の運動の複雑さ、人間活動の不確定性のために統計的予測法が有力な方法として用いられている。

現在、地方自治体が行っている大気汚染予報とか光化学スモッグ発生注意報はこの手法にもとづく予測結果をもとに出されている。

この統計的予測法ないしは計算法と対照的なのが、力学的ないしは物理的方法である。前者が実測資料に基づく経験方法であり、後者は少数個の普遍的基礎法則の上に構成される一般的な方法であるというように考えられている。

一般にこの二つの方法は相対立するもののように考えられているが、実はそうではない対象が物理現象であるならば理論的に両者の間の対応関係を導くことができるはずである。それは、力学的方法は物理法則の微分方程式による局所的表現から出発するのに対して、これを実用上便利な積分型のグリーン関数表示に直し、この関数を実測データより決定するのが統計的方法であると考えられるからである（水文系の場合には、理想的な流域についてすでに日野¹⁾によりこの関係は数式的に明らかにされている）。

* 正会員 工博 東京工業大学教授 工学部土木工学科

** (株) 数値解析研究所

*** 日本電気(株) 情報処理公共システム営業本部 システム部

(1) グリーン関数による拡散方程式の解

拡散の場合についてこの関係を説明する。まず、現象を単純化して一次元の拡散方程式を考える。

$$\frac{\partial C}{\partial t} + U \frac{\partial C}{\partial x} = K \frac{\partial^2 C}{\partial x^2} \dots\dots\dots (1)$$

この偏微分方程式の解は Green 関数 $G(t, x; \tau, \xi)$ を用いて次のように表すことができる。

$$C(t, x) = \iint G(t, x; \tau, \xi) s(\tau, \xi) d\tau d\xi \dots\dots (2)$$

ここに、

$$G(t, x; \tau, \xi) = \frac{1}{2\sqrt{\pi K(t-\tau)}} e^{-\frac{[(x-\xi)-U(t-\tau)]^2}{4K(t-\tau)}} \quad (t \geq \tau) \dots\dots\dots (2a)$$

U : 風速, K : 拡散係数, s : 排煙量

(2) 差分化による回帰型式の誘導

積分領域を間隔 $\delta\tau, \delta\xi$ の格子網で覆う。時間格子の番号は現在より過去に付けることとする。 $(\tau_\alpha, \xi_\beta), (t_i, x_j)$ における排煙量・濃度およびグリーン関数を $S_{\alpha, \beta}, C_{i, j}, G_{i, j, \alpha, \beta}$ とし、上式を差分近似すれば

$$C_{i, j} = \sum_{\alpha} \sum_{\beta} G_{i, j, \alpha, \beta} s_{\alpha, \beta} \dots\dots\dots (3)$$

ただし、面要素 $\delta\tau\delta\xi$ は単純化のため G の中に含めて考え、省略する。次に列ベクトル s_k, C_i, \tilde{C} を

$$\left. \begin{aligned} s_k &= [s_{k1} \ s_{k2} \ \dots \ s_{kn}]^T \\ C_i &= [C_{i1} \ C_{i2} \ \dots \ C_{in}]^T \\ \tilde{C} &= [C_1 \ C_2 \ \dots \ C_m]^T \end{aligned} \right\} \dots\dots\dots (4)$$

と定義する。また、 j 点のグリーン関数を $G_{i, j, \alpha, \beta} = G_{(\alpha-i)j\beta} \equiv G_{\alpha j\beta}$ ($\alpha = \alpha - i$) として、行ベクトル G_α および \tilde{G} を次のように定義する。

$$G_\alpha = [G_{\alpha 11} \ G_{\alpha 21} \ \dots \ G_{\alpha n1} \ G_{\alpha 1} \ \dots \ G_{\alpha n2} \ \dots \ G_{\alpha n\nu}]$$

$$\tilde{G} = [G_0 \ G_1 \ \dots \ G_\mu]$$

排煙量 $s_{\alpha, \beta}$ の影響範囲は時間的に $1 \leq \alpha < \mu$ であり、排煙源は x -格子 ($j=1 \sim N$) の十分内側にありその数は $\beta=1 \sim \nu$ とする。 x に関する格子点より n 個の点を選び、時間 t に関して m 回の C, s の側定値が得られれば、次の関係が成立する。

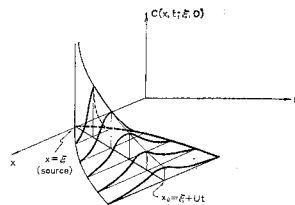


Fig. 1 Schematic representation of Green's function of one-dimensional diffusion from a source.

$$\begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_m \end{bmatrix} = \begin{bmatrix} G_0 & \dots & G_1 & \dots & G_\mu & 0 & \dots & 0 \\ 0 & G_0 & \dots & G_\mu & & & & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & G_0 & G_1 & \dots & G_\mu & \dots \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_\mu \\ s_{\mu+m-1} \end{bmatrix} \dots\dots\dots (5)$$

すなわち、

$$C_k = G_0 s_k + G_1 s_{k+1} + \dots + C_\mu s_{k+\mu} \dots\dots\dots (6)$$

一般に、排煙量はまったく不規則な変化をすることはなく、統計的規則性ないしは自己および相互相関をもって変化する。いま、 s_k の変化を簡単に一次のマルコフ過程¹⁵⁾

$$s_k = r s_{k-1} + r_k \dots\dots\dots (7)$$

(r_k : 乱数ベクトル) とすれば、濃度の予測は過去の濃度のデータに関して次のように表わされる。

$$C = \frac{1}{N} \sum_{k=1}^N r^k C_k + e \quad (e: \text{誤差ベクトル}) \dots\dots (8)$$

風速 U や拡散係数 K が一定ならば、三次元の場合には Green 関数を

$$G(x, y, z, t; \xi, \eta, \zeta, \tau) = \frac{1}{[4\pi K(t-\tau)]^{3/2}} \cdot \exp\left\{-\frac{[(x-\xi)-U(t-\tau)]^2 + [y-\eta]^2}{4K(t-\tau)}\right\} \times \left[e^{-\frac{(z-\zeta)^2}{4K(t-\tau)}} + e^{-\frac{(z+\zeta)^2}{4K(t-\tau)}}\right] \dots\dots\dots (9)$$

として、上の議論はそのまま成立する。

考え方を簡明に示すために二、三の制約をもうけたが、以上述べたことから過去の濃度データにより将来の濃度予測の回帰式が求められることを示すことができた。実際の大气汚染の場合には、拡散場は非均質な3次元場であり、風速・風向・大気安定度も入力項として入ってくるので式(8)、(9)の関係は単純ではないにしても、原理的には上に述べた考え方は成立し、統計的方法も物理的背景をもっているといえるであろう。

表一 大気汚染予測モデルの分類⁹⁾

分類	方法	モデル名
物理モデル	Physical model	数値シミュレーション (日野, APCs) ⁹⁾ (大西) (島貫)
	Simplified physical model (plume or puff model)	ブルーム (横山) ⁷⁾ ブルーム (林) 二層ブルーム (公害気象) 二層パフ (日立) ⁹⁾
	Simplified physical and statistical model	混合層 (高松—Least sq.) ⁹⁾ 多重ボックス (日立) ⁹⁾
非物理モデル	Control theory	集中定数系 (日野—Kalman filter) ¹⁾
	Statistical method	回帰分析 (東京都) 因子分析 (藤田)

(3) グリーン関数の実測データによる推定

なお、排煙濃度のみでなく排煙量に関してもデータが得られている場合には、直接的にグリーン関数を求めることができる。

$$\begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_m \end{bmatrix} = \begin{bmatrix} s_1^T & s_2^T & \dots & s_{\mu+1}^T & 0 & \dots & 0 \\ 0 & s_2^T & \dots & s_{\mu+1}^T & s_{\mu+2}^T & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & s_m^T & \dots & \dots & s_{\mu+m}^T & \dots \end{bmatrix} \begin{bmatrix} G_0^T \\ G_1^T \\ \vdots \\ G_\mu^T \end{bmatrix} \quad (10)$$

すなわち

$$\tilde{C} = S_* \tilde{G} \quad (11)$$

したがって、 $m > (\mu+1) \times \nu$ ならばグリーン関数は最小二乗法により次のように求められる。

$$\tilde{G} = [S_*^T S_*]^{-1} S_*^T \tilde{C} \quad (12)$$

3. 本論文の予測法の概説

本論文の汚染濃度予測法は、直接的には物理方程式によらない統計的予測法である（物理モデルとの内在的関連は前節に述べた）。統計的手法としては

- 因子分析}静的予測法
- 回帰分析}静的予測法
- カルマン・フィルター.....動的予測法

を組み合わせて用いる。この方法を **FRAK** (Factor analysis, Regression analysis & Kalman filter) 法と名付けたいと思う。

因子分析はデータの次元（地点数）を縮小するためであるが、後に触れるように回帰予測の精度の上からもプラスの点がある。カルマン・フィルターは予測と実測の差を最新の情報とみなして、これを積極的に利用するためである。本論文の方法のうち最も特徴的なものは、回帰式を各時刻ごとに別々に決定したこと（位相回帰（または周期回帰）モデルを考えたこと）である。

(1) 位相回帰モデル

さて、自然現象や社会現象は多くの場合にきわだった周期変化を伴っている。それは、これらが太陽エネルギーを活動源としており、かつ地球が自転公転という周期運動をしていることによる。したがって、式(8)の線型変換行列 r^K も一定不変と考えるよりは、一周期内の位相 T をパラメーターとして含むか、位相ごとに異なるとして取扱った方がよいと考えられる。

大気汚染の場合についていえば、汚染濃度 C の変化を記述するすべての初期条件・境界条件・入力-初期濃度の分布、境界濃度の分布、汚染物排出の位置と量、風速・風向の空間的分布、拡散能に影響する気象要素（例

えば、気温・日射）は時間的空間的に複雑に変化しておりこれらを正確に計測することは不可能に近い。しかし、これらの条件はまったくデータラメに変化するのではなく、1日の平均の変化があり、日々の条件はその平均曲線のまわりに不規則な（一般には平均値に比べて大きくはない）変化を行っている。

したがって、自然現象と人間の社会活動の日周期を考慮し、各時刻ごとに回帰関係式を仮定する。すなわち、回帰係数行列 A は1日の位相(時刻)ごとに異なるとする。

$$z(t) = A(T)x(t) + c(T) + e \quad (13)$$

ここに、 T は1日のうちの位相(時刻)を表わす。

$$T = \text{MOD}(t/T_0) \quad (T_0: \text{周期})$$

また、 $z(t)$ は予測すべき n 個の成分より成るベクトル、 $x(t)$ は現時点までに得られている m 個の成分からなる説明変数ベクトル、 e は誤差ベクトル、 $c(T)$ は回帰ベクトル。

ただし、本論文では実データではなく、これを変換し

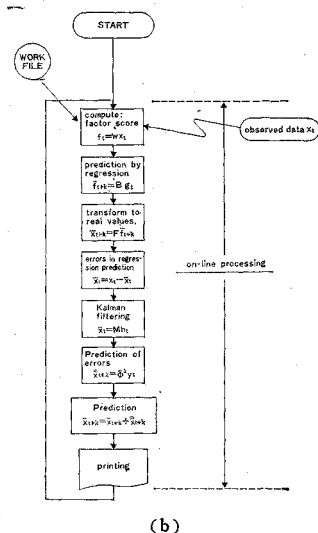
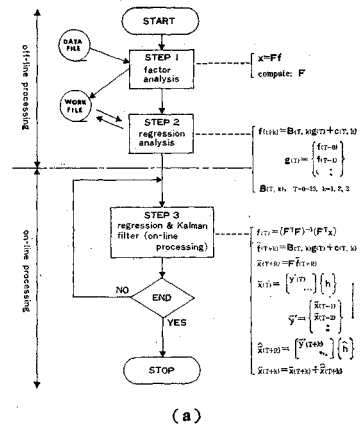


Fig. 2 Flow-chart of APMP prediction scheme by FRAK method.

て求めた因子評点時系列について位相回帰式を求める。

(2) システム構成

上述の統計的予測手法を組み合わせて構成された予測システム (APMS のサブシステム) は、図-2 (a), (b) のようである。すなわち、まず off-line システムとしてファイルされた測定データを用いて

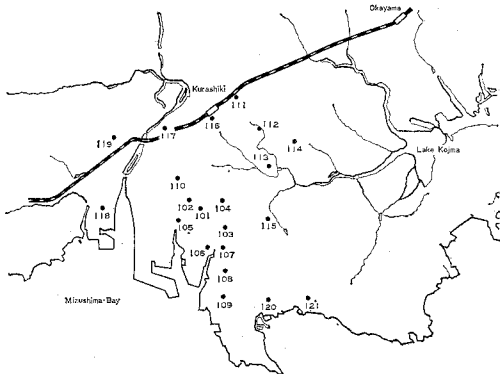
Step 1: 因子分析→因子パターン (負荷) 行列
因子評点時系列

Step 2: 位相回帰分析→位相回帰係数行列

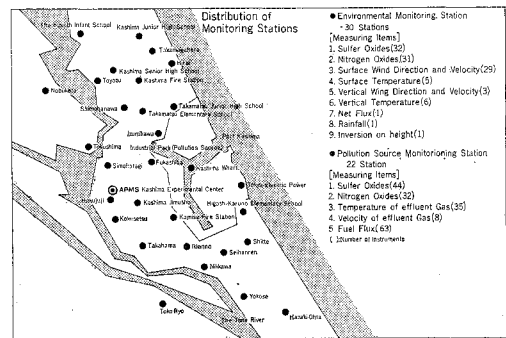
の二つの Step を行う。この Step がいわゆる training の段階で、1~2 か月間の測定データがあてられる。求められた因子パターン行列 F と位相回帰係数行列 B はファイルにストアされる。

これらの予備処理と準備の後、on-line による予測システムによる Step 3 に移る。

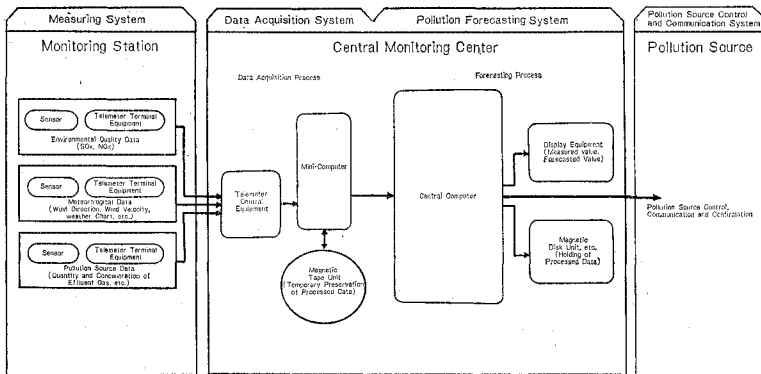
- Step 3
- ① 測定データ→因子評点時系列→位相回帰予測→予測濃度
 - ② 予測誤差→カルマン・フィルター→予測値補正



(a)



(b)



(c)

Fig. 3 (a) Locations of monitoring stations at Mizushima industrial area.
(b) Network of monitoring stations at Kashima heavy industrial area.
(c) Schematic illustration of Air Pollution Monitoring and Control System.

Step 1, 2 に要する machine time は、地点数 7・因子縮小 4・データ期間 30 日の場合、NEAC-2200/500 により約 40 分である。

Step 3 は地点数 7, 予測は 1.2・3 時間先の 3 種類の場合 2~3 分ほどで完了し、on-line 予測として条件を十分満たすものと思われる。

〔対象地域および期間〕 解析の対象としたのは、茨城県鹿島地区と岡山県水島地区である (図-3)。鹿島地区に 20 km×20 km の範囲に環境濃度 SO₂ (30 地点) NO_x (25 地点) および気象要素 (風向風速・気温・日射量等) の測定網がめぐらされ、中央局に刻々通報される。中央局ではミニコン (NEAC-M 4) 2 台によりデータの処理が行われ、中央コンピューター (NEAC-2200/500) により一時間ごとに 1.2・3 時間先の濃度予測が行われる。濃度予測は本論文の方法 (FRAK 法, APMS では通称「線型予測」と plume & puff model により行われる。このシステムの詳細については、文献 14), 15) を参されたい。

水島地区の予測は、現地で採録されたデータ・テープを鹿島地区の中央コンピューターに結んで再生実験的に

予測性の検討を行った。

4. 因子分析による前処理

汚染濃度の予測に先立って、因子分析により汚染状況の解析を行うことが望ましい。これは、予測を因子評点ではなく、生のデータについて行う場合にも対象地域の汚染の特性を把握し、予測性を高めるために必要である。

因子分析は正準因子分析法^{12),13)}によった。第 n 日の時刻 i 地点 j の濃度 $C_j^n(i)$ は、正規化され

$$x_j^n(i) = \frac{C_j^n(i) - \bar{C}_j}{\sqrt{(C_j^n(i) - \bar{C}_j)^2}} \dots\dots\dots(14)$$

の形で取り扱われている。

(a) 計測地点の共通性・独自性

同一時刻での各測定地点の濃度と代表地点の風速成分についての因子分析

$$\mathbf{x} = \mathbf{F}\mathbf{f} + \tilde{\mathbf{f}} \dots\dots\dots(15)$$

(ここに、 \mathbf{x} :濃度 C_j と風速成分 $U \cdot V$ から成るベクトル、 \mathbf{F} : 因子負荷行列、汚染濃度分布のパターンを表わす行列、 \mathbf{f} : 因子評点、各因子の時刻ごとの大きさを表わすベクトル、誤差ベクトル)を行った。結果のうちから例を表-2, 3 に示す。

(b) 因子評点による濃度変化の再現性

式(15)から因子負荷行列 \mathbf{F} が求まると、元のデータ \mathbf{x} に対応する因子評点時系列 $\mathbf{f}(t)$ が計算できる。

$$\mathbf{f}(t) = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{x} \dots\dots\dots(16)$$

この $\mathbf{f}(t)$ に \mathbf{F} を作用させて得られる $\bar{\mathbf{x}}(t)$ が、

$$\bar{\mathbf{x}}(t) = \mathbf{F}\mathbf{f}(t) \dots\dots\dots(16a)$$

表-2 SO₂ の因子負荷行列および独自性 (水島, 6月)

No.	地点番号	因子負荷行列						独自性
		第1因子	第2因子	第3因子	第4因子	第5因子	第6因子	
1	102	0.87	-0.24	0.07	-0.17	-0.03	-0.07	0.14
2	103	0.78	-0.17	-0.19	-0.01	-0.23	0.08	0.26
3	105	0.89	-0.33	-0.07	-0.01	0.07	0.02	0.08
4	109	0.60	-0.35	0.20	0.40	0.02	-0.02	0.31
5	111	0.90	0.13	0.25	-0.02	-0.07	0.02	0.10
6	112	0.93	0.25	-0.03	-0.01	0.07	0.01	0.06
7	113	0.88	0.18	-0.18	0.12	-0.07	-0.06	0.13

表-3 SO₂ の因子負荷行列および独自性 (鹿島, 3月)

No.	地点番号	因子負荷行列					独自性
		第1因子	第2因子	第3因子	第4因子	第5因子	
1	101	0.41	0.52	0.19	0.19	-0.06	0.49
2	102	0.20	0.25	0.32	-0.31	0.02	0.70
3	106	0.86	-0.15	-0.14	-0.04	0.05	0.21
4	107	0.32	-0.14	0.22	-0.27	-0.16	0.73
5	108	0.23	0.24	0.39	-0.01	0.18	0.70
6	109	0.62	0.42	-0.28	-0.06	-0.02	0.36
7	114	0.79	-0.22	0.20	0.11	-0.04	0.27

どれほど元の $\mathbf{x}(t)$ を再現しうるかを調べたのが 図-4, 5 である。

因子分析における独自性の低い水島地区の因子による再現性はかなり高く、鋭い濃度のピークも意外によく再現されている。また、地点数や因子数の多いほど再現性は高い。これに反し、独立性の高い地点の多い鹿島地区の因子による再現性は低い。前者は多煙源型、後者は集中少数煙源型地区であることを考えると興味深い。

(c) 因子分析 (空間分布) における日周期性考慮の効果

回帰式のみならず、因子分析の段階で汚染の日周期性を考慮した解析も行った。つまり、因子パターン行列 $\mathbf{F}(T)$ も因子評点 $\mathbf{f}(t, T)$ も位相 T の関数と考える。この場合は因子分析に用いるデータ数が 1/24 になるという不利な点が生じる。実際に因子評点による再現性を比較したが、各時刻をベタにした場合との差はみられなかった。

本システムとしてはこの2つを組み込みいずれかを使えるようになっているが、混乱を避けるため以下には回帰式の段階にのみ位相を考慮した方法について述べる。

5. 予測理論の現地実験による検討

(1) 予測の training data と test data

上述のような予備処理や検討を行ったのち、現地においてオン・ライン予測の実験に入った。予測計算には、最新の 20~50 日の data を training data として、因子分析と位相回帰分析 (Step 1・2) の操作を行い、これに続く 1 週間ほどを test (または check) data とした。結果は実験期間中 (50 年 11 月~51 年 9 月) ひと月ごとに月報として報告される (実験期間中は、他のモデルのテストや調整が行われるので、連日継続してオンライン予測を行っているのではない)。

(2) 回帰式の予測性

(a) 回帰式の決定

濃度の空間分布の因子が抽出されると、普通ならば濃度について行われる回帰式の決定を、濃度変化を説明する因子の評点 \mathbf{f} について求めることができる。すなわち、位相を考慮した因子レベルでの周期回帰式 (論文 3) p. 48, 式 (23))

$$\mathbf{f}(t+k) = \mathbf{B}(T, k) \mathbf{g}(t) \dots\dots\dots(17)$$

$$\bar{\mathbf{x}}(t+k) = \mathbf{F}\mathbf{f}(t+k) \dots\dots\dots(18)$$

を決定する。ここに、

$$\mathbf{g}(t) = [\mathbf{f}^T(t) \mathbf{f}^T(t-1) \mathbf{f}^T(t-2) \dots]^T \dots\dots(19)$$

表-4 は、時刻 $T=0$ 時の 1 時間先の予測式の回帰

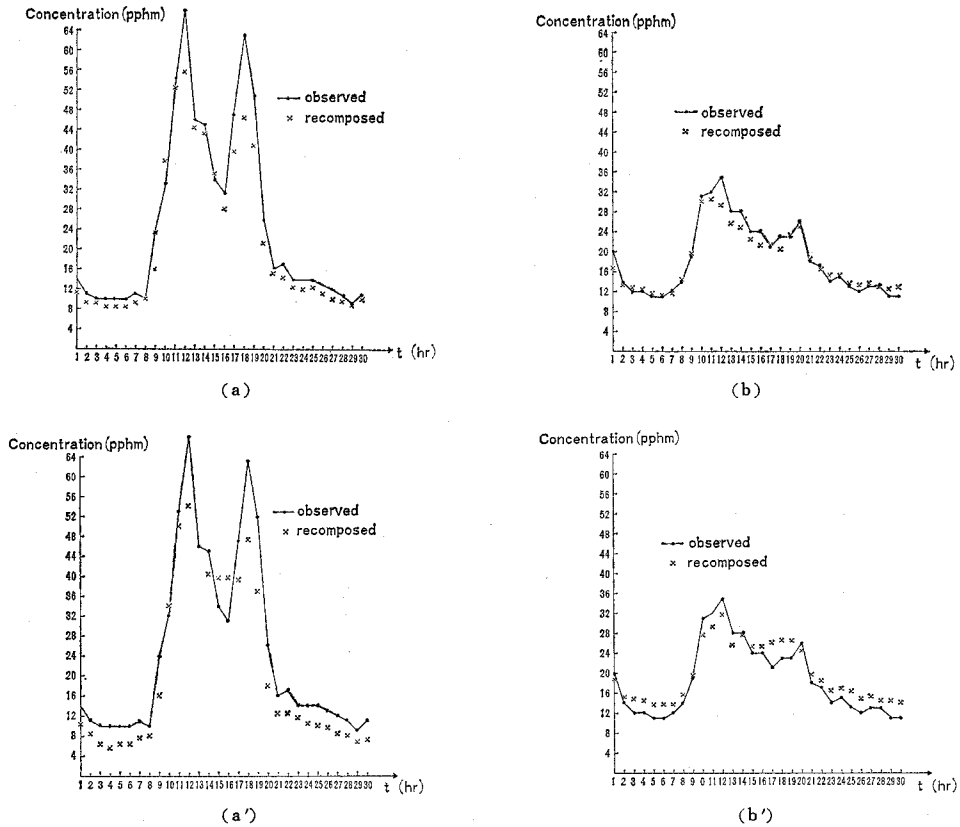


Fig. 4 (a), (b) Test of the reliability of factor analysis; number of variables=15 and number of factors=10; (a) Mizushima No. 102 station, (b) Mizushima No. 109 station.
 (a'), (b') Test of the reliability of factor analysis; number of variables=7, number of factors=5; (a') Mizushima No. 102, (b') Mizushima No. 109.

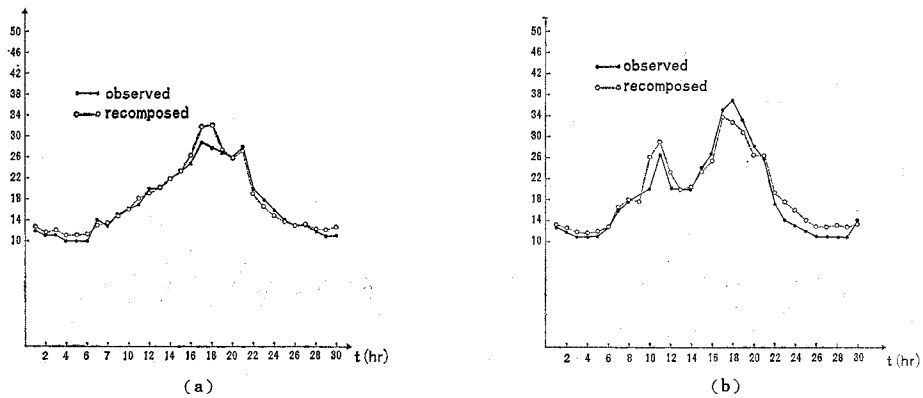


Fig. 5 Test of the reliability of factor analysis at Kashima area.

表-4 回帰行列 B の値 (鹿島 SO₂) (ただし, 時刻 T=0) B(0.1)g

説明因子 因子 No.	f_0					f_{-1}					f_{-2}				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	1.593	-0.117	-0.043	0.235	-0.087	-0.832	0.273	0.462	-0.059	0.069	0.019	-0.016	0.032	-0.024	-0.072
2	0.850	1.049	-0.291	0.032	0.109	-1.022	0.195	0.310	0.346	-0.270	0.090	-0.075	-0.166	-0.334	0.039
3	0.127	-0.256	1.040	0.158	0.010	-0.286	0.268	0.156	-0.013	-0.392	0.134	-0.155	-0.025	-0.047	0.085
4	-0.268	0.215	0.051	0.328	-0.357	0.612	-0.197	-0.136	0.402	0.139	-0.100	0.064	-0.077	-0.139	0.266
5	0.040	-0.097	0.114	0.019	0.808	-0.338	0.189	0.087	0.085	-0.298	0.324	-0.158	0.018	-0.006	0.034

行列 $B^T(0, 1)$ である。 $f \cdot g$ の構成には第1～第5因子を採り、説明成分としては $T=0, 23, 22$ の3時間前までの f を選んでいる。このような回帰行列 B が各予測時間 ($k=1, 2, 3$) について各時刻ごとに $24 \times 3 = 72$ 面、計算されファイルされる。

(b) 位相を考慮したことの効果

因子分析(因子行列の決定)には時刻つまり位相を考慮しないが、因子評点の時系列から回帰式を決定する場合には位相ごとに回帰行列を求めた。第1因子の j -step 前の値 $f_i(t-j)$ が予測にどれだけ寄与するかは、回帰行列 B の第 $(5 \times j + i)$ 列要素を見ればよい。例えば、 B の第1列目は現時点の第1因子評点 $f_1(0)$ の k ($=1$) 時間ステップ後の因子評点の予測値 $\hat{f}(t+k)$ への寄与の度合を示す。 $f_1(0)$ の第1～3因子予測値への寄与度の時間変動を表わしたのが、図-6である。

逆に、 B の行要素は因子評点予測への $f_i(j)$ ($i=1, 2, \dots, 5$) の寄与度を表わす。図-7は1時間先の第1因子への現在および過去の各因子の寄与度の時刻変化である。

これら2つの図より、それぞれの因子の重要さは時刻により大きく入れ代ること、つまり位相回帰の考え方が有効であることが裏付けられる。なお、図-8は第1因子スコアの変動時系列の例である。

(c) 回帰予測を実データではなく因子レベルで行う効果

実データを因子評点時系列に変換縮小することは、変数の次元を縮小し演算を楽にするが、一方、実データもっている情報の一部を汜過することになるので、実データのままである回帰予測より予測性が低下するのではないかと危惧された。同一のデータについて実データ(つまり因子縮小を行わない場合)と因子評点に変換した場合の予測性の比較を行ったところ、因子縮小をしたものの方が優れていることがわかった。これは、共通性の低い雑音の情報も汜過されるためと考えられる(図-15と図-15(A)とを対照のこと)。

(d) 回帰式に用いる過去のデータの時点数と予測性
ここでは、回帰式に $0, -1, -2$ の現在および過去のデータを用いた。さらにさかのぼった過去のデータを用いても予測精度の向上はみられず、多くの場合多少予測精度は低下した。これは、位相回帰式の中に間接的に長周期変動の考え方が入っていること、過去のデータ数を増すことは(メモリーの関係から)同時処理の地点数を減らすことになるので両者の効果が相殺されるためである。

(e) 位相回帰予測の予測精度

位相回帰式による予測と実測との比較例を 図-9, 10 に示す。また、このときの予測値と実測値の相関を 図-

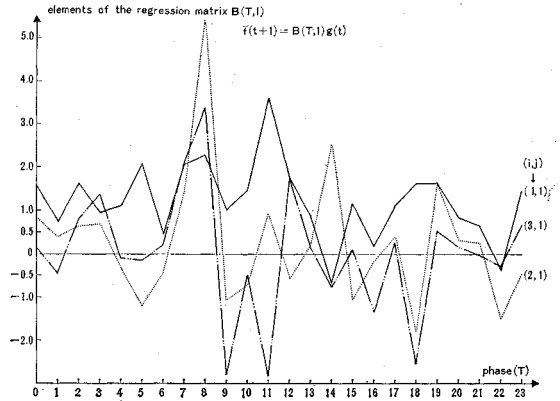


Fig. 6 Variation of the elements of the regression matrix $B(T, 1)$ for the one-step ahead prediction.

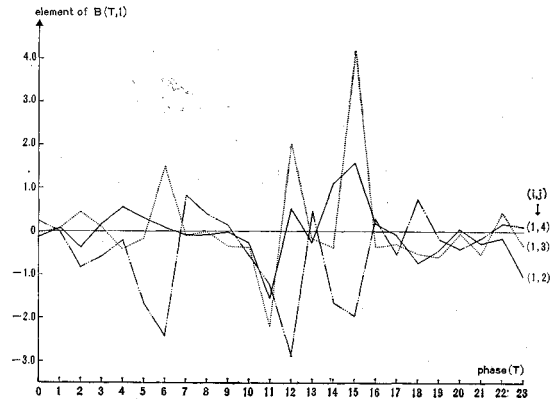


Fig. 7 Variation of the elements of the regression matrix $B(T, 1)$ for the one-step ahead prediction.

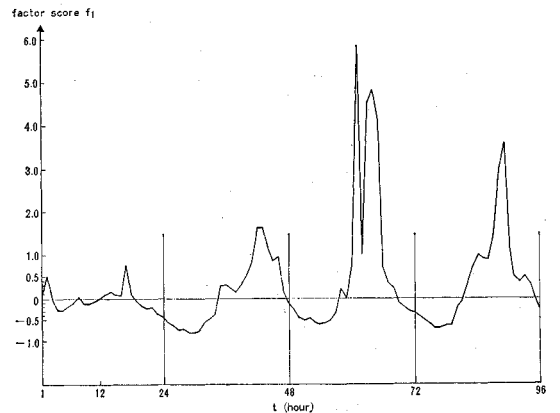


Fig. 8 Temporal change of a factor score f_1 .

11, 12 に示す。

結果はほぼ満足すべきものと考えられる。

(3) カルマン・フィルターの組み合わせによる予測性の向上

(a) 回帰予測残差

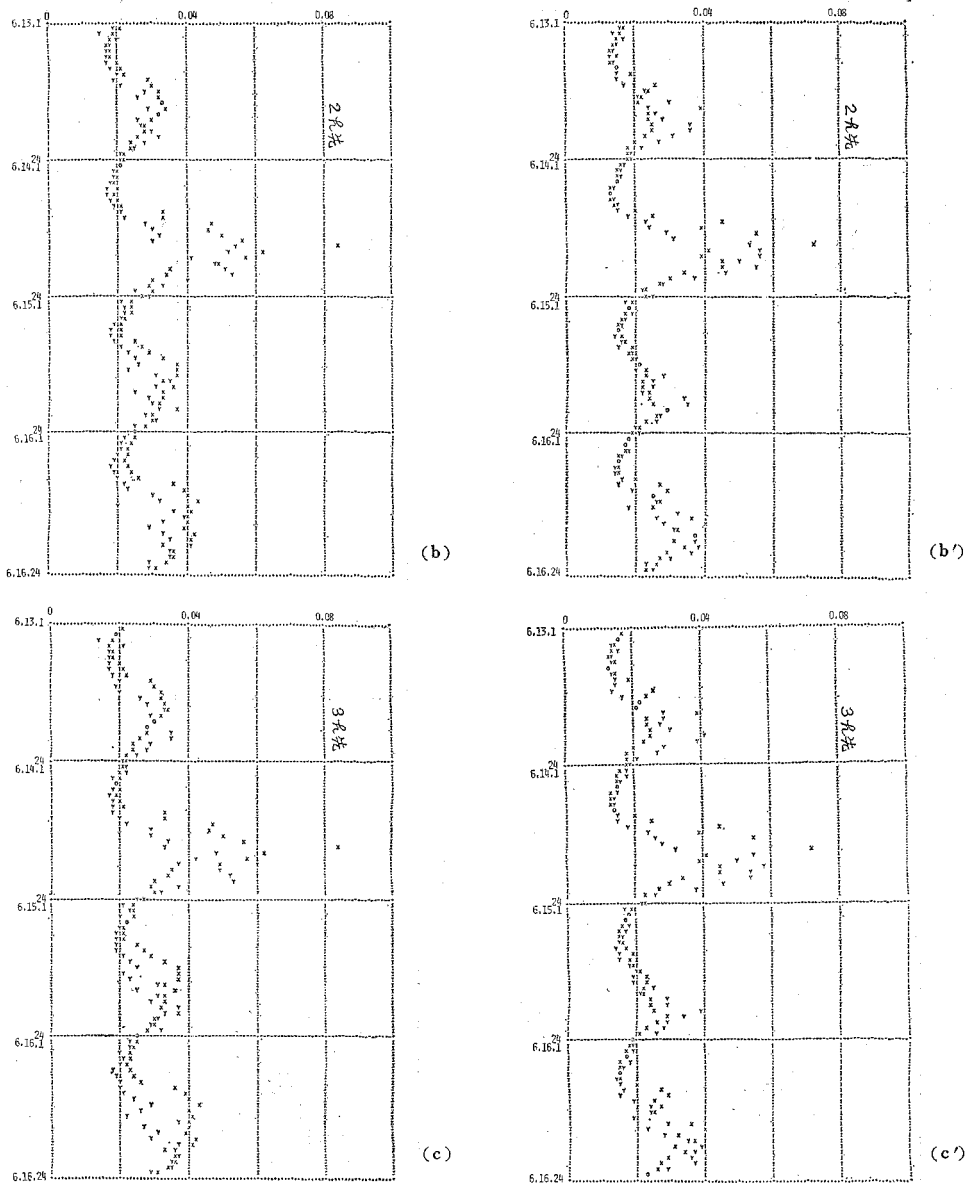
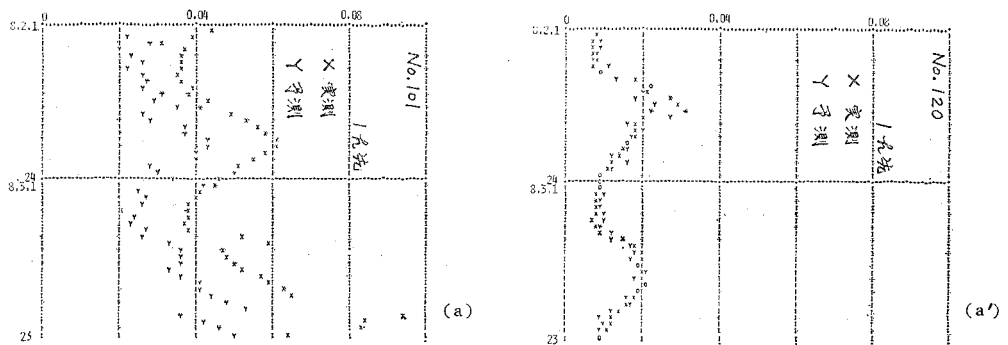


Fig. 9 Comparisons between the observed SO_2 concentration and the predicted concentration by the FRAK method. Eight monitoring stations are processed as one group. (a) 1 hr ahead prediction (b) 2 hr ahead prediction and (c) 3 hr ahead prediction.



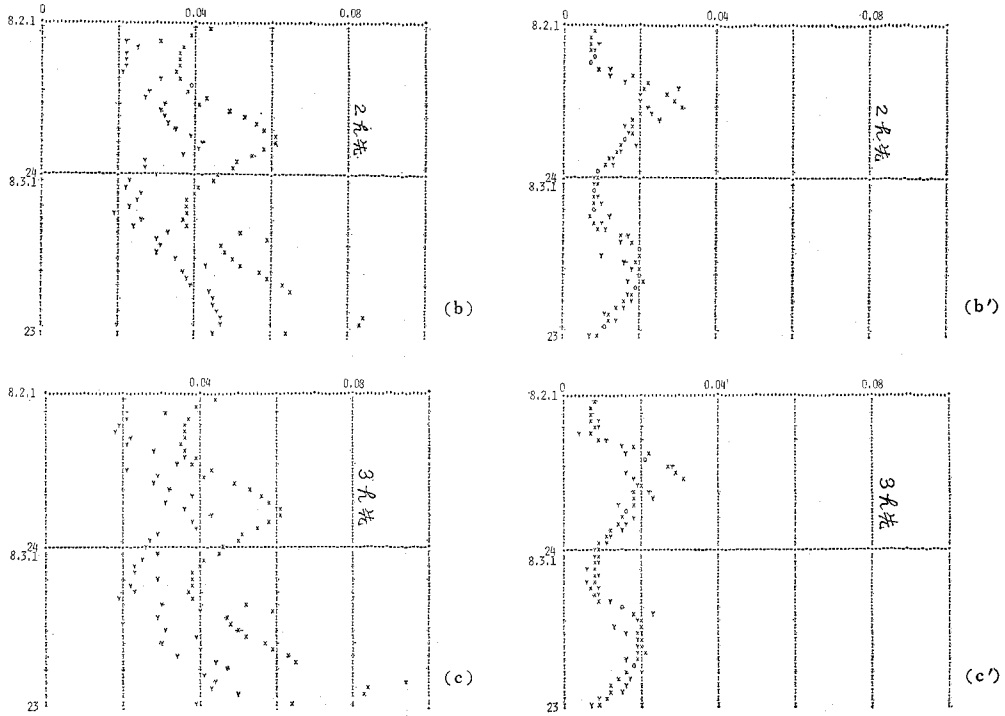
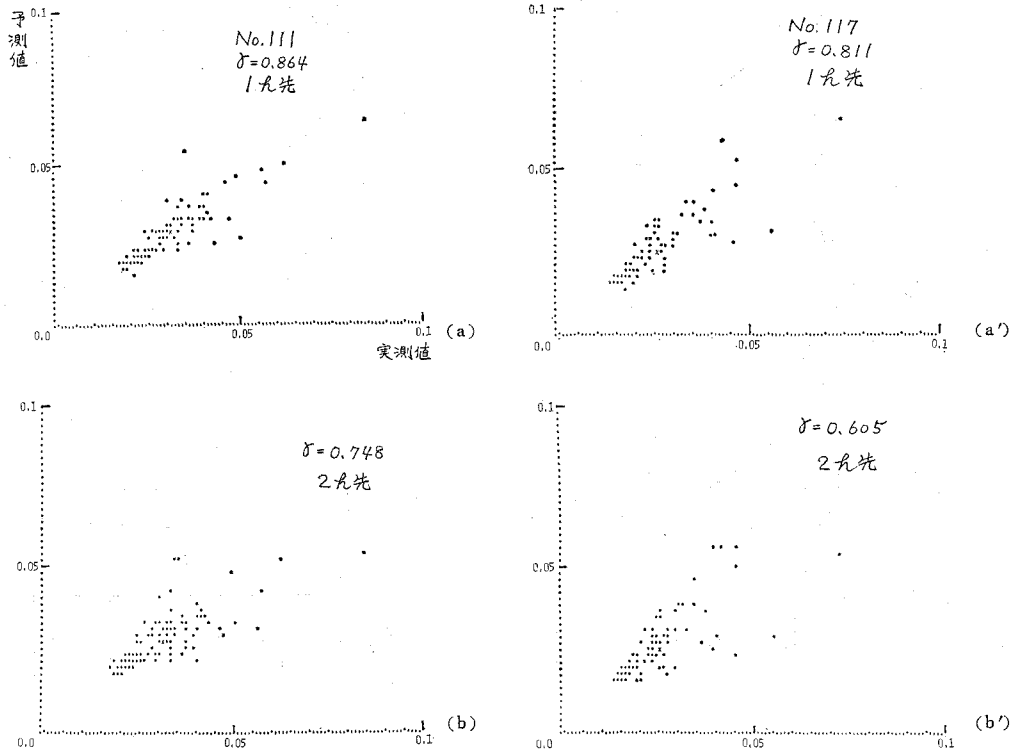


Fig. 10 Comparisons of the time series of observed SO₂ concentration with those predicted by the phase regression analysis only, at Kashima heavy industry area. (a) 1 hr ahead (b) 2 hr ahead and (c) 3 hr ahead predictions. Reference should be made to Fig. 15.



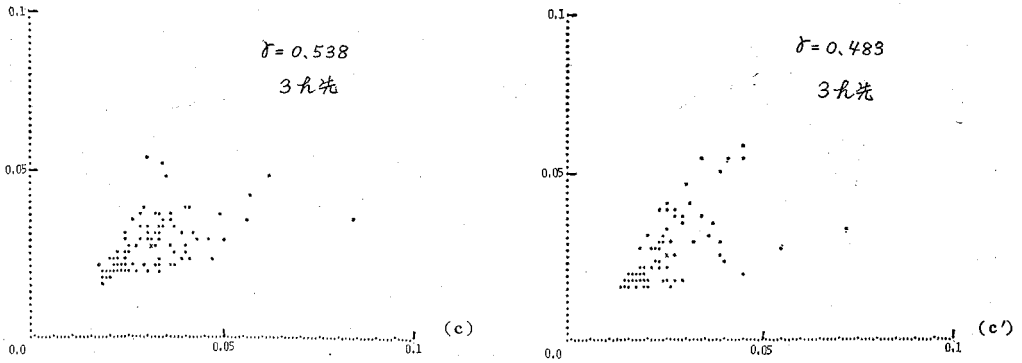


Fig. 11 Correlations between the observed and the predicted (FRAK) concentrations, at Mizushima area. (a) 1 hr ahead prediction, (b) 2 hr ahead and (c) 3 hr ahead predictions.

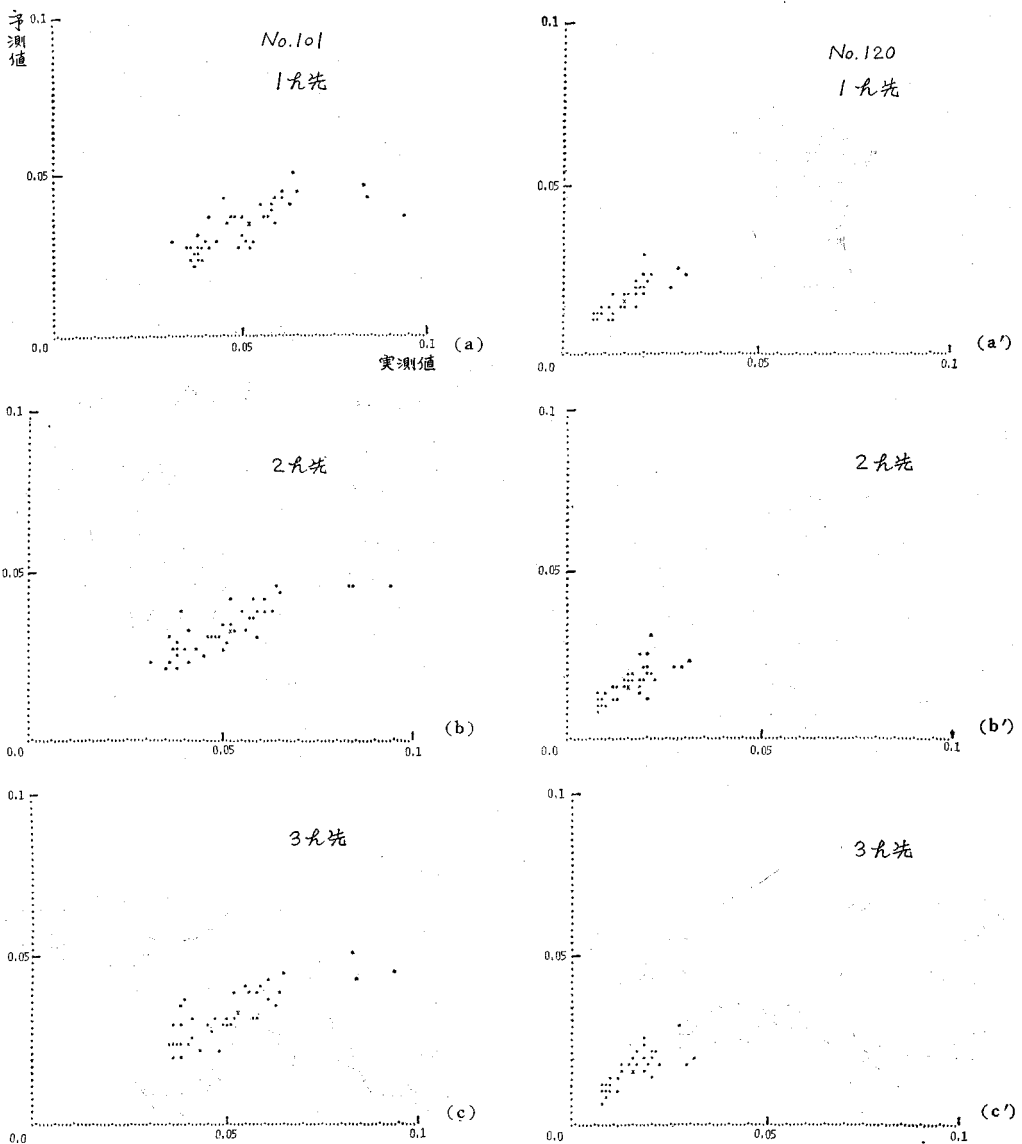


Fig. 12 Correlations between the observed and the predicted (phase regression alone) concentrations, at Kashima area. (a) 1 hr ahead, (b) 2 hr ahead and (c) 3 hr ahead predictions.

式 (28) を加えて次のようになる。

$$\hat{x}(t+k) = \bar{x}(t+k) + \hat{\tilde{x}}(t+k) \dots \dots \dots (30)$$

(b) システム方程式の項数 (過去のデータ数) と予測性

式 (21) においてどこまで過去のデータを用いればよいかを調べるために、次の 3 ケースについて予測精度

$$\alpha = \sqrt{\frac{1}{N} \sum (x_i - \hat{x}_i)^2 / N} \dots \dots \dots (29)$$

を比較した。

ケース 1: $y(t) = [\hat{x}_{t-1}^T]^T$

ケース 2: $y(t) = [\hat{x}_{t-1}^T, \hat{x}_{t-2}^T, \hat{x}_{t-3}^T]^T$

ケース 3: $y(t) = [\hat{x}_{t-1}^T, \hat{x}_{t-2}^T, \hat{x}_{t-3}^T, \hat{x}_{t-4}^T]^T$

結果は、表-5 のようで 2~3 ステップ前の比較的近い過去のデータを使用するだけで充分といえる。これは動的予測は主に周期の短い変動に対して有効であると

表-5 Kalman 予測におけるシステム方程式の項数と予測誤差の r.m.s.

ケース	1	2	3
予測ステップ			
1 時間先	5.94	7.75	7.26
2 時間先	4.57	5.68	5.70
3 時間先	3.40	3.57	3.52

いう考え方と一致する。

(c) 同定されたシステム係数の時間的変動

Kalman フィルターにより同定されたシステム方程式の係数がどう変化するかの一例が 図-13 である。この例では、係数はほぼゆるやかな時間変動をしているが、たまに急激な変化が生じることもあり、これは気象条件や排煙量の変化による濃度の急変に対応している。

(d) 実データと因子評点に対するフィルターリングの比較

著者らは最初 Kalman フィルターも因子評点レベルでの誤差ベクトル \hat{f} について行う考えであった。しかし、間もなくこれでは予測があまりよくなく、むしろ因子評点から実データに変換しその回帰予測誤差ベクトル \tilde{x} について Kalman フィルターをほどこせばよいことに気付いた (図-14)。因子予測による誤差分には、共通因子で説明できないものが残っているの、これは当然である。もし、因子レベルでフィルターリングを行うと独自性の高い地点の補正値は他地点の誤差補正に引きずられて、むしろ回帰予測結果を帳消しにするよう働くことが認められた。

(e) カルマン・フィルターによる補正を行った場合

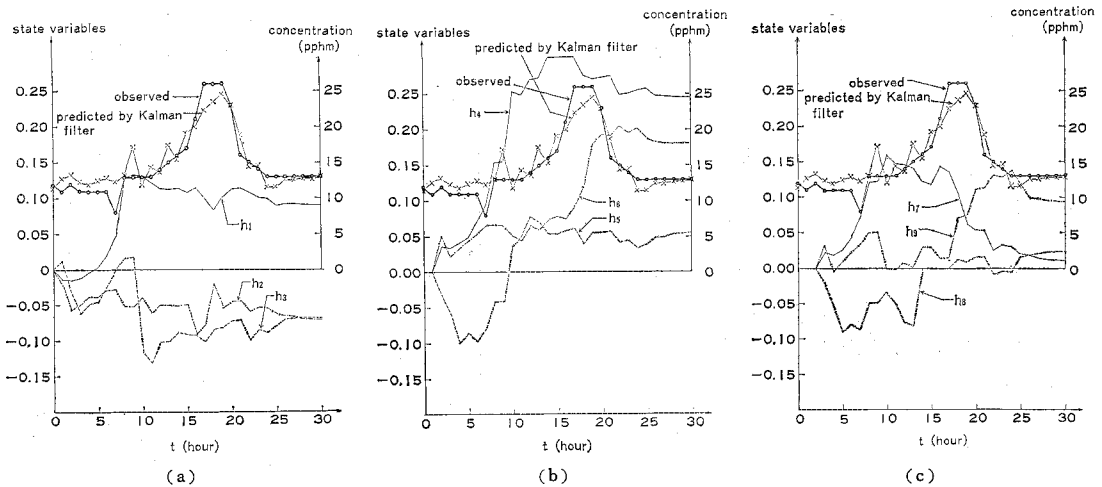


Fig. 13 (a), (b), (c) Temporal changes in the identified system parameters h_i .

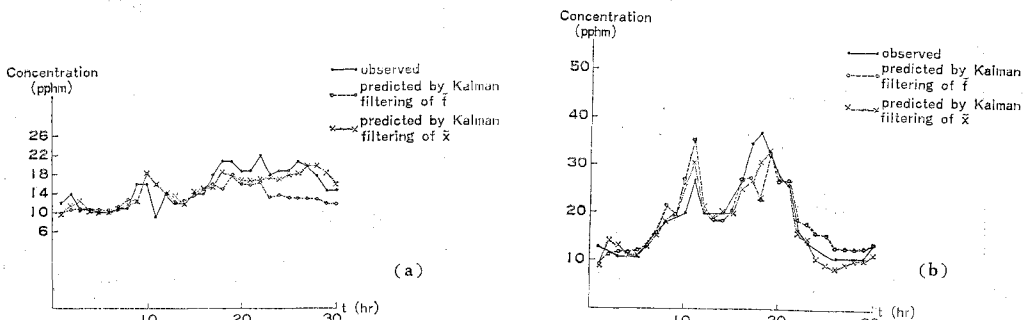


Fig. 14 The predictions by Kalman filtering on \hat{f} and \tilde{x} , compared with the observed values.

の予測精度

図-10 に示した 位相回帰予測に、さらにカルマン・フィルターによる補正を行った場合の予測結果を 図-15 に示す。予測値と実測値の相関を 図-16 および 表-6 にあげる。これらの図から、カルマンフィルターが位相

表-6 予測値と実値の相関係数 (1975 年 8 月)

局	1h 先	2h 先	3h 先	備考
101	0.877	0.816	0.745	
102	0.796	0.765	0.770	
104	0.060	0.134	0.228	濃度が低い地点
113	0.362	0.325	0.410	濃度が低い地点
118	0.720	0.659	0.691	
120	0.858	0.823	0.769	
123	0.823	0.744	0.712	
129	0.808	0.711	0.713	

回帰式による予測結果をさらに高めていることは明瞭である。予測と実測との相関がそれほど高くない地点(104, 113)は濃度そのものの低い地点であり、問題はない。したがって、現地における多数地点での種々の条件下のテストにより、本論文の大気汚染予測方法は(SO₂ 予測に関して)ほぼ満足すべきものと結論しうるのであろう。また、乱流現象の本質から考えてこれ以上の予測精度の向上も困難であろう。

結 論

本論文は、著者の一人(MH)がこれまで研究提案してきた多変量解析(静的方法)と最適制御理論(動的方法)を組み合わせた環境汚染濃度のオンライン予測法

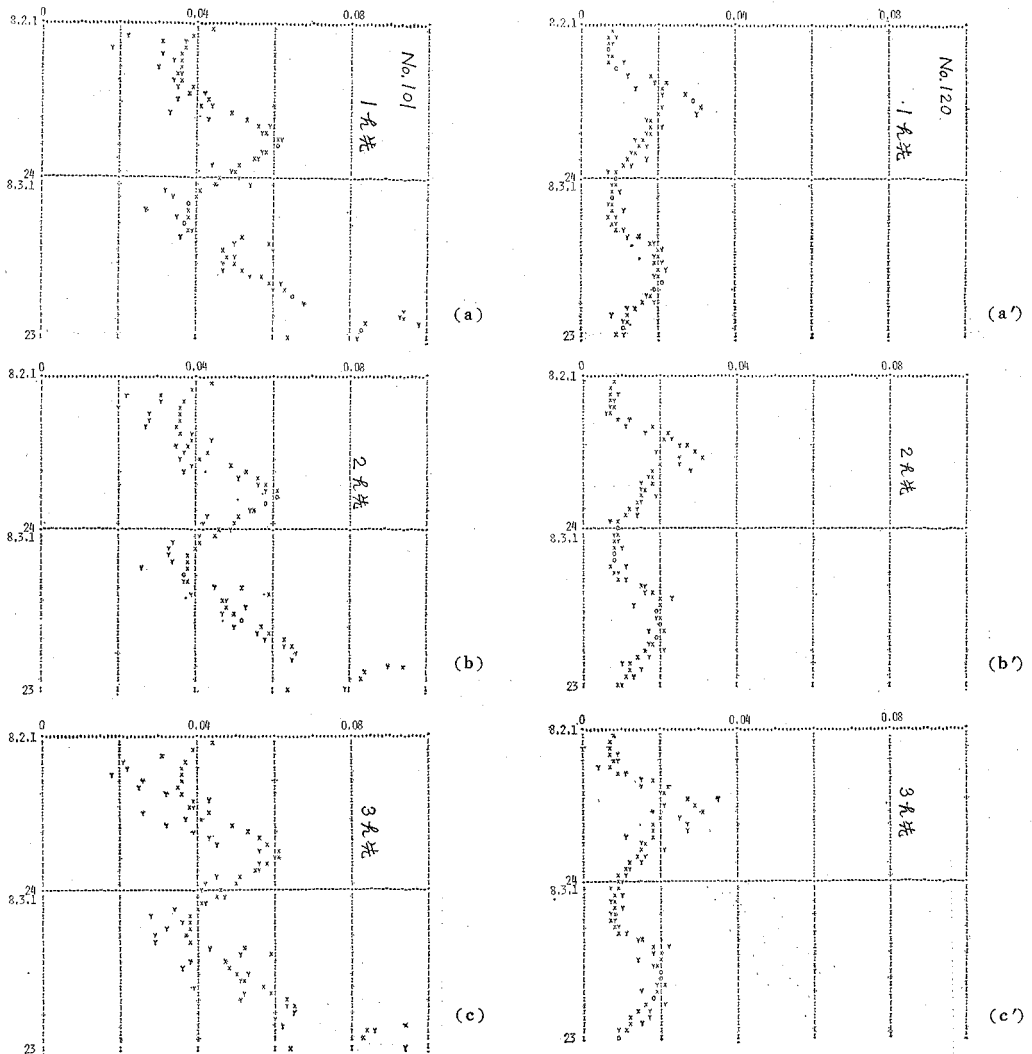
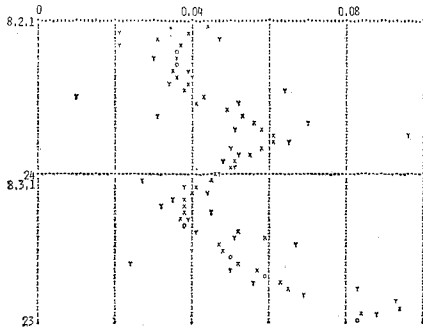
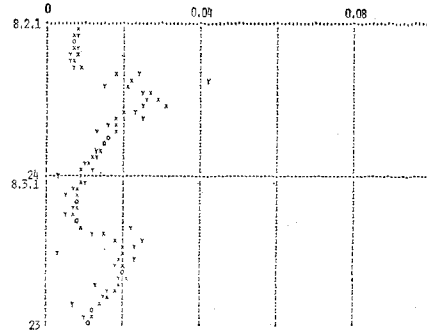


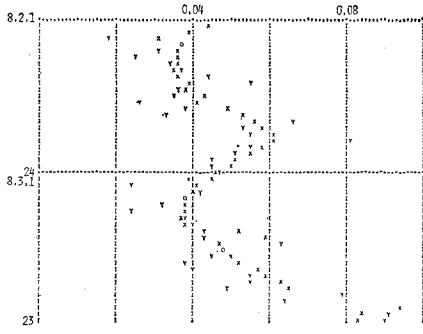
Fig. 15 Comparison between the observed and the predicted (phase regression and Kalman filter) concentrations. The predictability is remarkably increased (c.f. Fig. 10). (a) 1 hr ahead, (b) 2 hr ahead and (c) 3 hr ahead predictions.



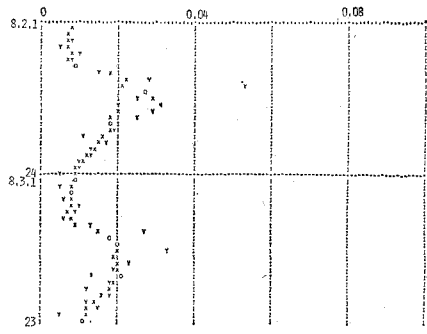
(a)



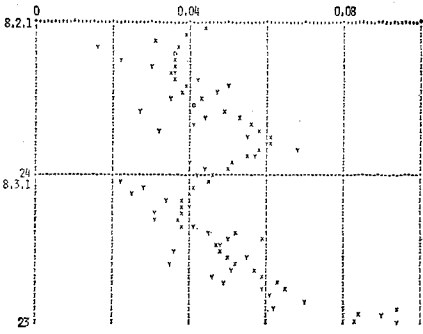
(a')



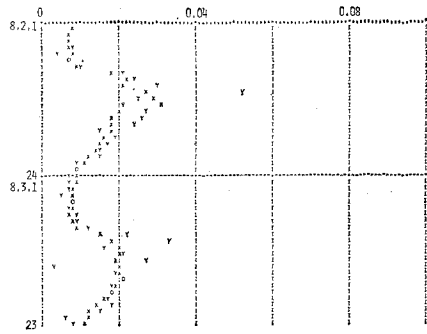
(b)



(b')

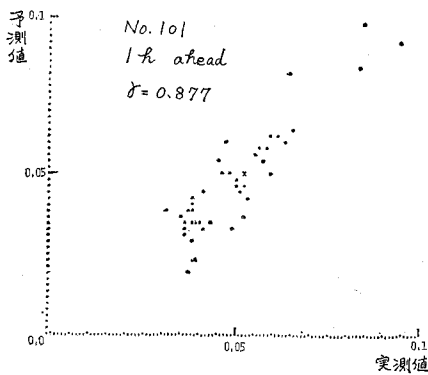


(c)

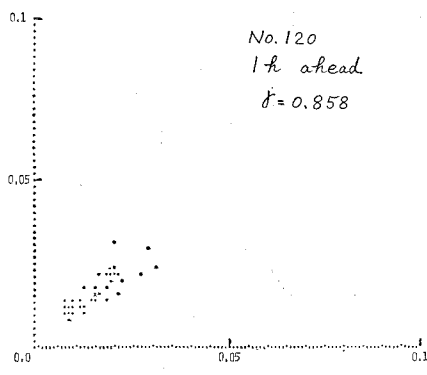


(c')

Fig. 15 A Prediction by the FRAK method without degeneration of number of factors.



(a)



(a')

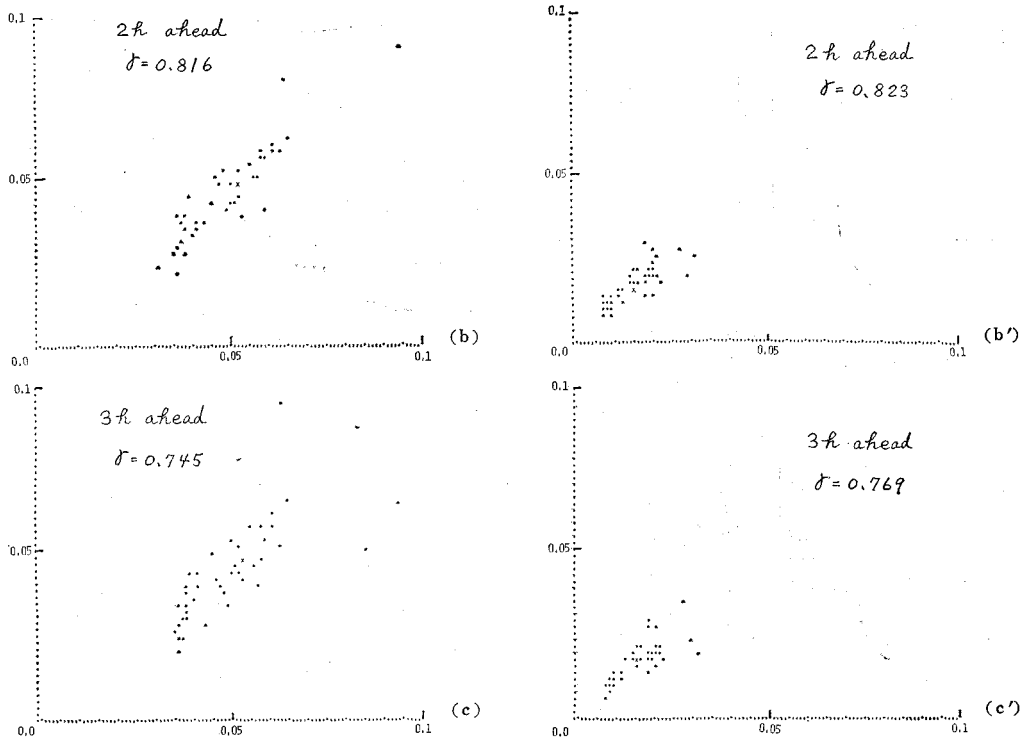


Fig. 16 Correlations between the observed and the predicted (FRAK) concentrations. (a) 1 hr ahead, (b) 2 hr ahead and (c) 3 hr ahead predictions.

(FRK 法あるいは KARFA 法と略称)を、茨城県鹿島地区および岡山県水島地区に設置された予測システム(環境・気象・煙源のモニター系と中央コンピューターによる予測系より成る)で運用した結果の報告である。これにより、次のように結論しうる。

(a) 回帰式型の統計予測モデルは単なるブラック・ボックス・モデルではなく、(差分化された)物理モデルの係数同定モデルとみなすことができる。

(b) 同時刻の気象条件・各測定点の濃度についての因子分析は、汚染状況の把握・変数の次元縮小のためのみでなく、予測のうえからも効果的である。

(c) 汚染濃度・気象条件の日周期変化を考慮し、(因子評点について)回帰式を各時刻ごとに決定することは、少ないデータで精度のよい予測をするのに有効である。多くの場合、この周期回帰モデルで十分の予測精度をあげうる。

(d) 実測濃度と予測濃度の誤差時系列を拡散系の状態についての最新の情報とみなして、Kalman フィルター理論により予測値の補正を行えば予測精度はさらに向上する。

(e) 本報告の方法による予測精度は1時間先の予測で実測との相関は0.8強、3時間先で0.7強である。

謝辞：本研究は財団法人 機械振興協会・新機械システムセンターのプロジェクトの一部として行われたも

のである。また、一部文部省科学研究費特定研究(1)(環境汚染制御、研究代表者 東大 大島康次郎教授)の補助を受けた。大島教授はじめ研究グループの諸先生、APMS 拡散研究グループ(代表者 坂上治郎 お茶の水大学教授)のメンバーの方々、(財)機械振興協会、通産省重工業課および公害防止指導課の諸氏および現地でのシステム・オペレーションを担当された日本電気(株)システム部の諸兄に深く謝意を表したい。

補 遺

(a) 本論文で用いた記号の意味

t : 時刻あるいは $4t$ きざみの時間ステップ数

T : 一日の時刻, 位相 $T = \text{mod}(t, 24)$

$x(t)$: 実データ(濃度・風速成分)ベクトル

$f(t)$: 因子評点ベクトル

F : 因子負荷(パターン)行列, $x(t) = Ff(t)$

W : 因子評点算出行列, $f(t) = Wx(t)$

$g(t)$: 過去の f ベクトルから成るベクトル,

$$g(t) = [f^T(t), f^T(t-1), \dots]^T$$

$\hat{y}(t)$: 過去の誤差ベクトルから成るベクトル,

$$\hat{y}(t) = [\hat{x}^T(t), \hat{x}^T(t-1), \dots]^T$$

$\bar{f}(t+k)$: 時刻 t における k ステップ先の回帰予測因子評点ベクトル

$\bar{x}(t+k)$: 時刻 t における k ステップ先の回帰予測実データ・ベクトル

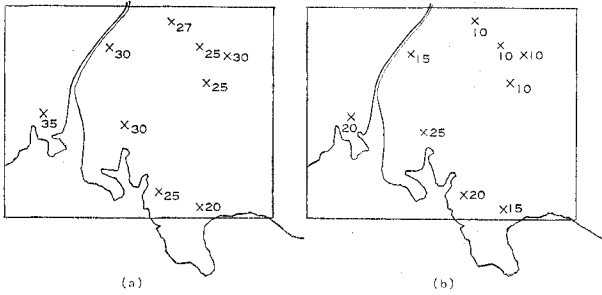


Fig. A1 SO₂ concentrations on nine monitoring stations (Mizushima).

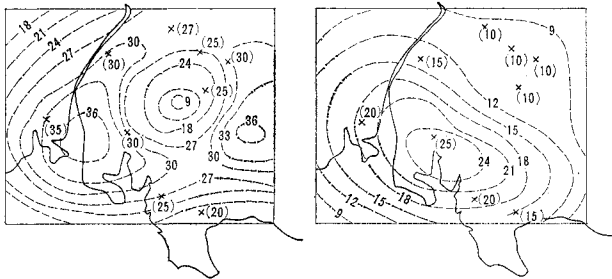


Fig. A The interpolated contour map of SO₂ concentration by the present method.

$B(T, k)$: 位相時刻 T において k ステップ先の予測を行う回帰行列

$\tilde{x}(t)$: 時刻 t における回帰予測誤差ベクトル,

$$\tilde{x}(t) = x(t) - \hat{x}(t)$$

Φ : 予測誤差に関する状態方程式の係数行列

h : 行列 Φ をカルマン・フィルター予測のために一列に並べ代えたベクトル

M : カルマン・フィルター予測のために誤差ベクトルより構成される (観測方程式) 行列.

~ : 誤差を意味する上つき添字

^ : 最適推定値を意味する上つき添字

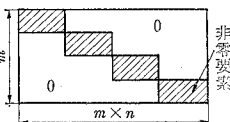
s : サブベクトルを意味する上つき添字

(b) 演算時間の短縮および記憶容量の縮小

オンライン処理演算 (とくにカルマン・フィルター部分) について, 次のような点を考慮して演算時間の短縮と記憶容量の縮小をはかった.

(i) 共分散行列 P は対称行列であるので, 全要素を記憶せず, 右上半分のみとした. これにより記憶容量を n^2 から $n(n+1)/2$ に減らし, したがって, 演算回数も激減できた.

(ii) 観測行列 M は非零要素が少なく, かつ共通部分も多いので, この無駄をはぶくことにより, $m^2 \times n$ の記憶容量を n に縮小した.



(c) 予測濃度分布図

統計制御理論的方法では, 観測点での濃度予測値が計算されるだけである. したがって, プロッターやブラウン管ディスプレイに自動的に等濃度線を描かせることが難しい. そこで, 各観測点の濃度 $C(x, y, z)$ を弾性平板のたわみと読み直し, これらのたわみは各観測点に作用する未知の荷重によると考える. グリーン関数を用いれば, この荷重は直ちに求められ, したがって, 任意点の濃度が計算できる²⁾.

グリーン関数としては, 四辺自由支持の場合の

$$G(x, y; \xi, \eta) = \frac{4}{\pi^4 ab} \sum_{m=1}^{m_0} \sum_{n=1}^{n_0} \left(\sin \frac{m\pi}{a} x \sin \frac{n\pi}{b} y \sin \frac{m\pi}{a} \xi \sin \frac{n\pi}{b} \eta \right) / \left(\frac{m^2}{a^2} + \frac{n^2}{b^2} \right)$$

を用いた. 辺の長さ a, b は観測領域を覆うよう十分大きくえらんだ. 図-A に例を示す.

参考文献

- 1) 日野幹雄・森 義一・吉川信二郎 (1974) : カルマンフィルターによる大気汚染の予測, 土木学会論文報告集, No. 224, 79-90.
- 2) 日野幹雄 (1974) : 因子分析・回帰分析およびカルマンフィルターによる環境濃度の予測, 文部省科学研究費一特定研究 (1) 「環境汚染制御」昭和 48 年度研究経過および研究報告書, 京大 榎木義一編, 137-144.
- 3) 日野幹雄 (1974) : カルマンフィルタと多変量解析を組み合わせた確率過程の予測法, 土木学会論文報告集, No. 228, 45-54.
- 4) 田島輝夫 : 大気汚染状況の因子分析による解析, 東工大土木工学科昭和 49 年度卒業論文.
- 5) Hino, M. (1968) : Computer experiment on smoke diffusion over a complicated topography, Atmospheric Environment, Vol. 2, 541-558.
- 6) 船橋誠寿・塩谷 真・春名公一・坂西 剛 (1972) : 大気汚染予測法, 日立評論, Vol. 54, No. 6.
- 7) 横山長之他 (1973) : APMS のパフ・ブルームモデル, 機械振興協会・APMS 報告書.
- 8) 高松武一郎・内藤正明 (1969) : 大気汚染の計算機制御, 計測と制御, 8 巻, 12 号.
- 9) 日野幹雄 (1973) : 大気汚染濃度の予測, 環境汚染制御に関する特別講演会, 講演要旨集, 19-35, 日本機械学会中国四国支部.
- 10) Kalman, R.E. (1960) : A new approach to linear filtering and prediction problems, J. Basic Eng., Trans ASME, Ser. D, Vol. 82, 35-45.
- 11) Hino, M. (1973) : Stochastic approach to linear and nonlinear runoff analysis, Proc. Int. Symp. on River Mechanics, Bangkok (also included in Tech. Report No. 12, Dept. Civil Eng., Tokyo Inst. of Tech.).
- 12) 浅野長一郎 (1971) : 因子分析法通論, 共立出版.
- 13) 芝 裕順 (1972) : 因子分析法, 東大出版会.
- 14) Sakagami, J., Hirose, M., Hino, M. & Yokoyama, O. (1976) : The air pollution forecast and control system, Proc. IFIP.
- 15) APMS 最終報告, 機械振興協会 (近刊, 1977).